

PATENT COOPERATION TREATY

PCT

COMMUNICATION OF
INTERNATIONAL APPLICATIONS

(PCT Article 20)

From the INTERNATIONAL BUREAU

To:

Commissioner
US Department of Commerce
United States Patent and Trademark
Office, PCT
2011 South Clark Place Room
CP2/5C24
Arlington, VA 22202
ETATS-UNIS D'AMERIQUE
in its capacity as designated Office

Date of mailing:

14 February 2002 (14.02.02)

The International Bureau transmits herewith copies of the international applications having the following international application numbers and international publication numbers:

International application no.:

PCT/CN00/00246

International publication no.:

The International Bureau of WIPO
34, chemin des Colombettes
1211 Geneva 20, Switzerland

Facsimile No.: (41-22) 740.14.35

Authorized officer:

J. Zahra

Telephone No.: (41-22) 338.83.38

PATENT COOPERATION TREATY

PCT

COMMUNICATION IN CASES FOR WHICH
NO OTHER FORM IS APPLICABLE

From the INTERNATIONAL BUREAU

To:

CCPIT PATENT AND TRADEMARK
LAW OFFICE
8th Floor
2 Fuchengmenwai St.
Beijing 100037
CHINE

Date of mailing (<i>day/month/year</i>) 14 February 2002 (14.02.02)	
Applicant's or agent's file reference IEE000019PCT	REPLY DUE see paragraph 1 below
International application No. PCT/CN00/00246	International filing date (<i>day/month/year</i>) 23 August 2000 (23.08.00)
Applicant INTEL CORPORATION	

1. ☐ REPLY DUE within _____ months/days from the above date of mailing
- ☐ NO REPLY DUE, however, see below
- ☒ IMPORTANT COMMUNICATION
- ☐ INFORMATION ONLY

2. COMMUNICATION:

The International Bureau regrets to inform the applicant that the above identified international application has not been published promptly after the expiration of 18 months from the priority date, as provided in PCT Article 21(2)(a).

International publication will now take place as soon as possible.

Meanwhile, the International Bureau will communicate a copy of the international application to each designated Offices, in accordance with PCT Article 20.

A copy of this notification is being sent to the receiving Office (RO/CN) and all designated Offices.

The International Bureau of WIPO 34, chemin des Colombettes 1211 Geneva 20, Switzerland	Authorized officer Anman QIU
Facsimile No. (41-22) 740.14.35	Telephone No. (41-22) 338.83.38



REQUEST

The undersigned requests that the present international application be processed according to the Patent Cooperation Treaty.

Receiving Office use only	
International Application No.	PCT/CN 00/00246
International Filing Date	23 AUG 2000 (23.08.00)
<div style="border: 1px solid black; padding: 2px; display: inline-block;"> RO/CN 中华人民共和国国家知识产权局 PCT International Application </div>	
Name of receiving Office and "PCT International Application"	
Applicant's or agent's file reference (if desired) (12 characters maximum)	IEE000019PCT

Box No. I	TITLE OF INVENTION	A METHOD AND APPARATUS FOR CONCEPT-BASED SEARCHING ACROSS A NETWORK	
Box No. II	APPLICANT		
Name and address: (Family name followed by given name; for a legal entity, full official designation. The address must include postal code and name of country. The country of the address indicated in this Box is the applicant's State (that is, country) of residence if no State of residence is indicated below.) INTEL CORPORATION 2200 Mission College Boulevard Santa Clara, California 95052 U. S. A.		<input type="checkbox"/> This person is also inventor. Telephone No. Facsimile No. Teleprinter No.	
State (that is, country) of nationality: US		State (that is, country) of residence: US	
This person is applicant for the purposes of: <input type="checkbox"/> all designated States <input checked="" type="checkbox"/> all designated States except the United States of America <input type="checkbox"/> the United States of America only <input type="checkbox"/> the States indicated in the Supplemental Box			
Box No. III	FURTHER APPLICANT(S) AND/OR (FURTHER) INVENTOR(S)		
Name and address: (Family name followed by given name; for a legal entity, full official designation. The address must include postal code and name of country. The country of the address indicated in this Box is the applicant's State (that is, country) of residence if no State of residence is indicated below.) INTEL CHINA LTD. Beijing Kerry Center 6/F, North Tower, No. 1 Guanghai Road, Chaoyang District Beijing 100020, CHINA		This person is: <input checked="" type="checkbox"/> applicant only <input type="checkbox"/> applicant and inventor <input type="checkbox"/> inventor only (If this check-box is marked, do not fill in below.)	
State (that is, country) of nationality: CN		State (that is, country) of residence: CN	
This person is applicant for the purposes of: <input type="checkbox"/> all designated States <input checked="" type="checkbox"/> all designated States except the United States of America <input type="checkbox"/> the United States of America only <input type="checkbox"/> the States indicated in the Supplemental Box			
<input type="checkbox"/> Further applicants and/or (further) inventors are indicated on a continuation sheet.			
Box No. IV	AGENT OR COMMON REPRESENTATIVE; OR ADDRESS FOR CORRESPONDENCE		
The person identified below is hereby/has been appointed to act on behalf of the applicant(s) before the competent International Authorities as:		<input checked="" type="checkbox"/> agent <input type="checkbox"/> common representative	
Name and address: (Family name followed by given name; for a legal entity, full official designation. The address must include postal code and name of country.) ⁹ CCPIT Patent and Trademark Law Office CCPIT PATENT AND TRADEMARK LAW OFFICE 8th Floor, 2 Fuchengmenwai St., Beijing 100037, CHINA		Telephone No. (86-10)68511076 Facsimile No. (86-10)68587610 (86-10)68587611 Teleprinter No.	
<input type="checkbox"/> Address for correspondence: Mark this check-box where no agent or common representative is/has been appointed and the space above is used instead to indicate a special address to which correspondence should be sent.			

Continuation of Box No. III FURTHER APPLICANT(S) AND/OR (FURTHER) INVENTOR(S)	
<i>If none of the following sub-boxes is used, this sheet should not be included in the request.</i>	
<p>Name and address: (Family name followed by given name; for a legal entity, full official designation. The address must include postal code and name of country. The country of the address indicated in this Box is the applicant's State (that is, country) of residence if no State of residence is indicated below.)</p> <p style="text-align: center; margin: 10px 0;">ZHOU, JOE F.</p> <p>Unite 4008 Capital Mansion, No. 6 Xin Yuan Nan Road, Chaoyang District, Beijing 100004, CHINA</p>	<p>This person is:</p> <p><input type="checkbox"/> applicant only</p> <p><input checked="" type="checkbox"/> applicant and inventor</p> <p><input type="checkbox"/> inventor only (If this check-box is marked, do not fill in below.)</p>
State (that is, country) of nationality: CN	State (that is, country) of residence: CN
<p>This person is applicant for the purposes of: <input type="checkbox"/> all designated States <input type="checkbox"/> all designated States except the United States of America <input checked="" type="checkbox"/> the United States of America only <input type="checkbox"/> the States indicated in the Supplemental Box</p>	
<p>Name and address: (Family name followed by given name; for a legal entity, full official designation. The address must include postal code and name of country. The country of the address indicated in this Box is the applicant's State (that is, country) of residence if no State of residence is indicated below.)</p> <p style="text-align: center; margin: 10px 0;">LIU, Weiquan</p> <p>Yu zhong Xi Li 27-1204, Xicheng District, Beijing 100029, CHINA</p>	<p>This person is:</p> <p><input type="checkbox"/> applicant only</p> <p><input checked="" type="checkbox"/> applicant and inventor</p> <p><input type="checkbox"/> inventor only (If this check-box is marked, do not fill in below.)</p>
State (that is, country) of nationality: CN	State (that is, country) of residence: CN
<p>This person is applicant for the purposes of: <input type="checkbox"/> all designated States <input type="checkbox"/> all designated States except the United States of America <input checked="" type="checkbox"/> the United States of America only <input type="checkbox"/> the States indicated in the Supplemental Box</p>	
<p>Name and address: (Family name followed by given name; for a legal entity, full official designation. The address must include postal code and name of country. The country of the address indicated in this Box is the applicant's State (that is, country) of residence if no State of residence is indicated below.)</p>	<p>This person is:</p> <p><input type="checkbox"/> applicant only</p> <p><input type="checkbox"/> applicant and inventor</p> <p><input type="checkbox"/> inventor only (If this check-box is marked, do not fill in below.)</p>
State (that is, country) of nationality:	State (that is, country) of residence:
<p>This person is applicant for the purposes of: <input type="checkbox"/> all designated States <input type="checkbox"/> all designated States except the United States of America <input type="checkbox"/> the United States of America only <input type="checkbox"/> the States indicated in the Supplemental Box</p>	
<p>Name and address: (Family name followed by given name; for a legal entity, full official designation. The address must include postal code and name of country. The country of the address indicated in this Box is the applicant's State (that is, country) of residence if no State of residence is indicated below.)</p>	<p>This person is:</p> <p><input type="checkbox"/> applicant only</p> <p><input type="checkbox"/> applicant and inventor</p> <p><input type="checkbox"/> inventor only (If this check-box is marked, do not fill in below.)</p>
State (that is, country) of nationality:	State (that is, country) of residence:
<p>This person is applicant for the purposes of: <input type="checkbox"/> all designated States <input type="checkbox"/> all designated States except the United States of America <input type="checkbox"/> the United States of America only <input type="checkbox"/> the States indicated in the Supplemental Box</p>	

☐ Further applicants and/or (further) inventors are indicated on another continuation sheet.

Box No. V DESIGNATION OF STATES

The following designations are made under Rule 4.9(a) (mark the applicable checkboxes; at least one must be marked):

Regional Patent

- ☒ AP ARIPO Patent: GH Ghana, GM Gambia, KE Kenya, LS Lesotho, MW Malawi, SD Sudan, SL Sierra Leone, SZ Swaziland, TZ United Republic of Tanzania, UG Uganda, ZW Zimbabwe, and any other State which is a Contracting State of the Harare Protocol and of the PCT
- ☒ EA Eurasian Patent: AM Armenia, AZ Azerbaijan, BY Belarus, KG Kyrgyzstan, KZ Kazakhstan, MD Republic of Moldova, RU Russian Federation, TJ Tajikistan, TM Turkmenistan, and any other State which is a Contracting State of the Eurasian Patent Convention and of the PCT
- ☒ EP European Patent: AT Austria, BE Belgium, CH and LI Switzerland and Liechtenstein, CY Cyprus, DE Germany, DK Denmark, ES Spain, FI Finland, FR France, GB United Kingdom, GR Greece, IE Ireland, IT Italy, LU Luxembourg, MC Monaco, NL Netherlands, PT Portugal, SE Sweden, and any other State which is a Contracting State of the European Patent Convention and of the PCT
- ☒ OA OAPI Patent: BF Burkina Faso, BJ Benin, CF Central African Republic, CG Congo, CI Côte d'Ivoire, CM Cameroon, GA Gabon, GN Guinea, GW Guinea-Bissau, ML Mali, MR Mauritania, NE Niger, SN Senegal, TD Chad, TG Togo, and any other State which is a member State of OAPI and a Contracting State of the PCT (if other kind of protection or treatment desired, specify on dotted line)

National Patent (if other kind of protection or treatment desired, specify on dotted line):

- | | |
|--|--|
| <input checked="" type="checkbox"/> AE United Arab Emirates | <input checked="" type="checkbox"/> LR Liberia |
| <input checked="" type="checkbox"/> AL Albania | <input checked="" type="checkbox"/> LS Lesotho |
| <input checked="" type="checkbox"/> AM Armenia | <input checked="" type="checkbox"/> LT Lithuania |
| <input checked="" type="checkbox"/> AT Austria | <input checked="" type="checkbox"/> LU Luxembourg |
| <input checked="" type="checkbox"/> AU Australia | <input checked="" type="checkbox"/> LV Latvia |
| <input checked="" type="checkbox"/> AZ Azerbaijan | <input checked="" type="checkbox"/> MA Morocco |
| <input checked="" type="checkbox"/> BA Bosnia and Herzegovina | <input checked="" type="checkbox"/> MD Republic of Moldova |
| <input checked="" type="checkbox"/> BB Barbados | <input checked="" type="checkbox"/> MG Madagascar |
| <input checked="" type="checkbox"/> BG Bulgaria | <input checked="" type="checkbox"/> MK The former Yugoslav Republic of Macedonia |
| <input checked="" type="checkbox"/> BR Brazil | <input checked="" type="checkbox"/> MN Mongolia |
| <input checked="" type="checkbox"/> BY Belarus | <input checked="" type="checkbox"/> MW Malawi |
| <input checked="" type="checkbox"/> CA Canada | <input checked="" type="checkbox"/> MX Mexico |
| <input checked="" type="checkbox"/> CH and LI Switzerland and Liechtenstein | <input checked="" type="checkbox"/> NO Norway |
| <input checked="" type="checkbox"/> CN China | <input checked="" type="checkbox"/> NZ New Zealand |
| <input checked="" type="checkbox"/> CR Costa Rica | <input checked="" type="checkbox"/> PL Poland |
| <input checked="" type="checkbox"/> CU Cuba | <input checked="" type="checkbox"/> PT Portugal |
| <input checked="" type="checkbox"/> CZ Czech Republic | <input checked="" type="checkbox"/> RO Romania |
| <input checked="" type="checkbox"/> DE Germany | <input checked="" type="checkbox"/> RU Russian Federation |
| <input checked="" type="checkbox"/> DK Denmark | <input checked="" type="checkbox"/> SD Sudan |
| <input checked="" type="checkbox"/> DM Dominica | <input checked="" type="checkbox"/> SE Sweden |
| <input checked="" type="checkbox"/> EE Estonia | <input checked="" type="checkbox"/> SG Singapore |
| <input checked="" type="checkbox"/> ES Spain | <input checked="" type="checkbox"/> SI Slovenia |
| <input checked="" type="checkbox"/> FI Finland | <input checked="" type="checkbox"/> SK Slovakia |
| <input checked="" type="checkbox"/> GB United Kingdom | <input checked="" type="checkbox"/> SL Sierra Leone |
| <input checked="" type="checkbox"/> GD Grenada | <input checked="" type="checkbox"/> TJ Tajikistan |
| <input checked="" type="checkbox"/> GE Georgia | <input checked="" type="checkbox"/> TM Turkmenistan |
| <input checked="" type="checkbox"/> GH Ghana | <input checked="" type="checkbox"/> TR Turkey |
| <input checked="" type="checkbox"/> GM Gambia | <input checked="" type="checkbox"/> TT Trinidad and Tobago |
| <input checked="" type="checkbox"/> HR Croatia | <input checked="" type="checkbox"/> TZ United Republic of Tanzania |
| <input checked="" type="checkbox"/> HU Hungary | <input checked="" type="checkbox"/> UA Ukraine |
| <input checked="" type="checkbox"/> ID Indonesia | <input checked="" type="checkbox"/> UG Uganda |
| <input checked="" type="checkbox"/> IL Israel | <input checked="" type="checkbox"/> US United States of America |
| <input checked="" type="checkbox"/> IN India | |
| <input checked="" type="checkbox"/> IS Iceland | |
| <input checked="" type="checkbox"/> JP Japan | <input checked="" type="checkbox"/> UZ Uzbekistan |
| <input checked="" type="checkbox"/> KE Kenya | <input checked="" type="checkbox"/> VN Viet Nam |
| <input checked="" type="checkbox"/> KG Kyrgyzstan | <input checked="" type="checkbox"/> YU Yugoslavia |
| <input checked="" type="checkbox"/> KP Democratic People's Republic of Korea | <input checked="" type="checkbox"/> ZA South Africa |
| | <input checked="" type="checkbox"/> ZW Zimbabwe |
| <input checked="" type="checkbox"/> KR Republic of Korea | |
| <input checked="" type="checkbox"/> KZ Kazakhstan | |
| <input checked="" type="checkbox"/> LC Saint Lucia | |
| <input checked="" type="checkbox"/> LK Sri Lanka | |

Check-boxes reserved for designating States which have become party to the PCT after issuance of this sheet:

- ☐
- ☐

Precautionary Designation Statement: In addition to the designations made above, the applicant also makes under Rule 4.9(b) all other designations which would be permitted under the PCT except any designation(s) indicated in the Supplemental Box as being excluded from the scope of this statement. The applicant declares that those additional designations are subject to confirmation and that any designation which is not confirmed before the expiration of 15 months from the priority date is to be regarded as withdrawn by the applicant at the expiration of that time limit. (Confirmation (including fees) must reach the receiving Office within the 15-month time limit.)

Sheet No. 4

Box No. VI PRIORITY CLAIMS		<input type="checkbox"/> Further priority claims are indicated in the Supplemental Box.		
Filing date of earlier application (day/month/year)	Number of earlier application	Where earlier application is:		
		national application: country	regional application: * regional Office	international application: receiving Office
item (1)				
item (2)				
item (3)				

☐ The receiving Office is requested to prepare and transmit to the International Bureau a certified copy of the earlier application(s) (only if the earlier application was filed with the Office which for the purposes of the present international application is the receiving Office) identified above as item(s):

* Where the earlier application is an ARIPO application, it is mandatory to indicate in the Supplemental Box at least one country party to the Paris Convention for the Protection of Industrial Property for which that earlier application was filed (Rule 4.10(b)(ii)). See Supplemental Box.

Box No. VII INTERNATIONAL SEARCHING AUTHORITY

Choice of International Searching Authority (ISA) (if two or more International Searching Authorities are competent to carry out the international search, indicate the Authority chosen; the two-letter code may be used):

ISA /

Request to use results of earlier search; reference to that search (if an earlier search has been carried out by or requested from the International Searching Authority):

Date (day/month/year)

Number

Country (or regional Office)

Box No. VIII CHECK LIST; LANGUAGE OF FILING

This international application contains the following number of sheets:

request : 4
 description (excluding sequence listing part) : 15
 claims : 6
 abstract : 1
 drawings : 7
 sequence listing part of description : _____

Total number of sheets : 33

This international application is accompanied by the item(s) marked below:

1. ☒ fee calculation sheet
2. ☐ separate signed power of attorney
3. ☐ copy of general power of attorney; reference number, if any:
4. ☐ statement explaining lack of signature
5. ☐ priority document(s) identified in Box No. VI as item(s):
6. ☐ translation of international application into (language):
7. ☐ separate indications concerning deposited microorganism or other biological material
8. ☐ nucleotide and/or amino acid sequence listing in computer readable form
9. ☐ other (specify):

Figure of the drawings which should accompany the abstract:

Fig. 1

Language of filing of the international application:

English

Box No. IX SIGNATURE OF APPLICANT OR AGENT

Next to each signature, indicate the name of the person signing and the capacity in which the person signs (if such information is not obvious from reading the request).



For receiving Office use only

1. Date of actual receipt of the purported international application: 23 AUG 2000 (23. 08. 00)	2. Drawings: <input type="checkbox"/> received: <input type="checkbox"/> not received:
3. Corrected date of actual receipt due to later but timely received papers or drawings completing the purported international application:	
4. Date of timely receipt of the required corrections under PCT Article 11(2):	
5. International Searching Authority (if two or more are competent): ISA /	6. <input type="checkbox"/> Transmittal of search copy delayed until search fee is paid.

For International Bureau use only

Date of receipt of the record copy by the International Bureau:

19 SEPTEMBER 2000

(19. 09. 00)

PATENT COOPERATION TREATY

PCT

REC'D 08 JUN 2001

WIPO

PCT

INTERNATIONAL SEARCH REPORT

(PCT Article 18 and Rule 43 and 44)

Applicant's or agent's file reference IEE000019PCT	FOR FURTHER ACTION see Notification of Transmittal of International Search Report (Form PCT/ISA/220) as well as, where applicable, item 5 below.	
International application No. PCT/CN00/00246	International filing date (day/month/year) 23.Aug 2000(23.08.00)	(Earliest)Priority date (day/month/year)
Applicant INTEL CORPORATION et al		

This international search report has been prepared by this International Searching Authority and is transmitted to the applicant according to Article 18. A copy is being transmitted to the International Bureau.

This international search report consists of a total of 2 sheets.

☐ It is also accompanied by a copy of each prior art document cited in this report.

1. Basis of the report

a. With regard to the **language**, the international search was carried out on the basis of the international application in the language in which it was filed, unless otherwise indicated under this item.

☐ the international search was carried out on the basis of a translation of the international application furnished to this Authority (Rule 23.1(b)).

b. With regard to any **nucleotide and /or amino acid sequence** disclosed in the international application, the international search was carried out on the basis of the sequence listing:

- ☐ contained in the international application in written form.
- ☐ filed together with the international application in computer readable form.
- ☐ furnished subsequently to this Authority in written form.
- ☐ furnished subsequently to this Authority in computer readable form.
- ☐ the statement that the subsequently furnished written sequence listing does not go beyond the disclosure in the international application as filed has been furnished.
- ☐ the statement that the information recorded in computer readable form is identical to the written sequence listing has been furnished.

2. ☐ **Certain claims were found unsearchable** (See Box I).

3. ☐ **Unity of invention is lacking** (See Box II)

4. With regard to the **title**,

- ☒ the text is approved as submitted by the applicant.
- ☐ the text has been established by this Authority to read as follows:

5. With regard to the **abstract**,

- ☒ the text is approved as submitted by the applicant.
- ☐ the text has been established, according to Rule 38.2(b), by this Authority as it appears in Box III. The applicant may, within one month from the date of mailing of this international search report, submit comments to this Authority.

6. The figure of the **drawings** to be published with the abstract is Figure No. 1

- ☒ as suggested by the applicant.
- ☐ because the applicant failed to suggest a figure.
- ☐ because this figure better characterizes the invention.

☐ None of the figures.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN00/00246

A. CLASSIFICATION OF SUBJECT MATTER

G06F17/30

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC7 G06F15/16 G06F17/30 G06F7/00

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	JP-A-2000057151(HITACHI LTD)25.Feb 2000(25.02.00) The whole document	1-30
A	US-A-6014665 (CULLISS G)11. Jan 2000(11.01.00) The whole document	1-30
A	WO-A-0026762 (SEARCHGENIE.COM INC)11.May 2000(11.05.00) The whole document	1-30

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:

- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier application or patent but published on or after the international filing date
- "L" document which may throw doubts on priority claim (S) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search
12.May 2001(12.05.01)

Date of mailing of the international search report

24 MAY 2001 12.05.01

Name and mailing address of the ISA/CN
6 Xitucheng Rd., Jimen Bridge, Haidian District,
100088 Beijing, China
Facsimile No. 86-10-62019451

Authorized officer

Wang Tao

Telephone No. 86-10-62093049



A METHOD AND APPARATUS FOR CONCEPT-BASED SEARCHING ACROSS A NETWORK

FIELD

5 The invention relates to searching across a network. More specifically, the invention relates to a method and apparatus for a concept-based search of web sites within a network.

*different
version
via W002/
27 541*

BACKGROUND

10 With the advent of the Internet and the World Wide Web (WWW), the vast amount of information available across the different web sites on the Internet continues to grow as the number of networks and servers connected thereto continue to increase on a global scale. Accordingly, the availability of this information on the Internet enables users to obtain such information quickly. However, locating such information on the
15 Internet can sometimes be difficult because of the vast numbers of websites available.

 Currently, particular web sites on the WWW include search mechanisms termed "search engines" that have been developed in order to assist the users of the WWW in locating specific information therein. In particular, software programs, termed crawlers or spiders, automatically traverse, on a periodic basis, the WWW through the different
20 web sites located thereon based on hyperlinks among the different web sites. When these software programs locate a web site whose information is not stored in the local database for this particular search engine, such programs transmit keywords and other relevant information concerning the new web site back to the web site executing these programs. Accordingly, this information about the new websites is stored into the local database for
25 the search engine website.

 Moreover, because of the dynamic nature of the WWW, these software programs also update such relevant information concerning a web site whose information is already

stored in the search engine's local database. Accordingly, these search engine web sites store these search terms in a local database with other search terms from other websites. In operation, a user of these search engines enters a search query into the search engine and based on this local database, the search engine locates web pages on different web sites whose search terms include the search query and retrieves portions of these web pages. The search engine then displays these portions of the web pages to the user, thereby allowing the user to select from among the different web pages.

Additionally, another type of search engine for the WWW performs a search among the search engine results described above. In particular, these type of search engines perform a search among the individual search engines as described above and combines the results from these individual search engines into a combined results, thereby allowing the user to obtain results from across multiple search engines for a single search query.

BRIEF DESCRIPTION OF THE DRAWINGS

Embodiments of the invention may be best understood by referring to the following description and accompanying drawings which illustrate such embodiments. In the drawings:

Figure 1 is block diagram illustrating a system that incorporates embodiments of the present invention;

Figure 2 illustrates one of servers 104*a-i* that include embodiments of the present invention;

Figure 3 is a flowchart illustrating a method for performing an expansive search for data information across servers 104 coupled to network 102, according to embodiments of the present invention;

Figure 4 illustrates a web page for entering a search term or phrase by clients

Figure 5 illustrates an example of a portion of a network of conceptual terms, according to embodiments of the present invention;

Figure 6 illustrates a method for the formation of concept networks 206, according to embodiments of the present invention; and

5 **Figure 7** illustrates a display of the output of concordance tool 218 used by a data analyst to identify terms related to a given anchor term.

Detailed Description

10 In the following description, numerous specific details are set forth to provide a thorough understanding of the invention. However, it is understood that the invention may be practiced without these specific details. In other instances, well-known structures and techniques have not been shown in detail in order not to obscure embodiments of the present invention.

15 **Figure 1** is block diagram illustrating a system that incorporates embodiments of the present invention. In particular, Figure 1 illustrates system 100 that includes servers 104*a-i* and clients 106*a-i*, which are coupled together through network 102. Servers 104*a-i* can include one to any number of such servers in system 100. In an embodiment, servers 104*a-i* are host computers, which store different types of data information
20 including one to a number of different web sites with each web site have one to a number of web pages thereon. Additionally, clients 106*a-i* can include one to any number of such clients in system 100. Clients 106*a-i* are client computers that have the network capability to connect to servers 104*a-i* through network 102.

25 While different embodiments of the present invention could include different types of networks, in an embodiment, network 102 is the Internet that includes the WWW. However, embodiments of the present invention could include other types of networks. For example, in one embodiment, network 102 is a local area network (LAN).

In another embodiment, network 106 is a wide area network (WAN). Further, network 106 can be a combination of different networks that provide communication among servers 104*a-i* and clients 106*a-i*. In operation for a given client, such as client 106*a*, connecting to a given server, such as server 104*a*, client 106*a* connects to server 104*a* through network 102 and is able to download and upload different web pages from and to different websites residing on server 104*a*.

Figure 2 illustrates one of servers 104*a-i* (hereinafter “servers 104”) that include embodiments of the present invention. As shown, servers 104 include processing unit 204, memory 202 and website 220. Additionally, website 220 includes database 208, concept networks 206, query reformulation unit 210, summarizer unit 212, search engine 214, network generation unit 216 and concordance tool 218, which are coupled together as shown. The following description of servers 104 is in terms of a particular web site on servers 104. However, the capability and/or functionality of servers 104 are not so limited, as servers 104 can service and/or host a number of web sites, which clients 106*a-i* (hereinafter “clients 106”) can access across network 102.

Database 208 and concept networks 206 are associated with website 220, which is stored on servers 104. In particular, website 220 acts as a search engine that enables clients 106 to search web pages on the local server, which stores this website, as well as remote servers coupled to network 102. In particular, website 220 is a search engine that allows the users thereof to enter search terms on a given web page and have this search engine return the results of a search. Additionally, concept networks 206 are defined to include one to any number of networks, whose creation is described in more detail below in conjunction with Figure 6. In an embodiment, each network within concept networks 206 is related to a given subject matter domain. For example, one network relates to computers, while another network in concept networks 206 relates to shoes.

Moreover, servers 104 include query reformation unit 210, search engine unit 212, summarizer 214, network generation unit 216 and concordance tool 218, which, in

one embodiment, are software programs that reside in memory 202 and processing unit 204 during their execution on processing unit 204. Query reformation unit 210, search engine unit 212, summarizer 214, network generation unit 216 and concordance tool 218 will be described in more detail below. In conjunction with system 100 of Figure 1 and the more detailed illustration of servers 104 of Figure 2, a method of operation of system 100 in accordance with embodiments of the present invention will now be described with reference to the flow chart shown in Figure 3.

In particular, **Figure 3** is a flowchart illustrating a method for performing an expansive search for data information, such as documents within web pages, across servers 104 coupled to network 102, according to embodiments of the present invention. Figure 3 illustrates method 300 that commences with the receipt of an initial search query, at process block 302. **Figure 4** illustrates a web page for entering a search term or phrase by clients 106. In particular, Figure 4 illustrates web page 402 that includes box 404 for entering the search term or phrase and go button 406 to initialize the search. Once client 106 enters a search term or phrase into box 404 and presses go button 406, web page 402 is transmitted back to server 104, which receives this initial search query, as shown at process block 302.

Upon receiving this initial search query, server 104 reformulates this query, at process block 304. In particular, server 104 searches a network within concept networks 206 to find terms related to the initial search term or phrase. **Figure 5** illustrates an example of a portion of a network within concept networks 206, according to embodiments of the present invention. As shown, Figure 5 includes network portion 500 that includes company's name 502, the company's subsidiaries 504, the company's products 506 and the company's stock ticker name 508, which are linked together. For example, network portion 500 could include the term "INTEL®" (company's name 502). Accordingly, company's products 506 could include the term "PENTIUM®" as one of the

name of the processors that INTEL[®] produces. Moreover, network portion 500, in this example, would include "INTC" as stock ticker name 508.

Moreover as shown in Figure 5, each pair of related terms in the network is connected by a link that specifies the semantic relationship between the related terms. As illustrated, products 506 are products of company's name 502. Additionally, subsidiaries 504 are members of company's name 502. Network portion 500 also includes a synonym relationship for denoting a similar name or phrase to a given name. In particular, stock ticker name 508 is a synonym of company's name 502. Embodiments of the present invention are not limited to the examples of semantic relationships illustrated in network portion 500. For example, another type of semantic relationship that can be incorporated into embodiments of the present invention is a canonical one. In particular, a canonical relationship is defined as the simplest form of a given word. For example, the simplest form of a telephone is a communication device.

"Hyponym of" is another example of a semantic relationship. A hyponym relationship denotes a relationship of a subcategory to a more general class. An illustration of a hyponym is a chair being a subcategory of furniture. Moreover, another semantic relationship is a hypernym one, which is a relationship of a general class to a subcategory. For example, automobile is a hypernym of a car. Further, another semantic relationship between the terms within the network is a "part" relationship. For example a bicycle tire is part of a bicycle. Accordingly, each link with concept networks 206 includes a description of the semantic relationship between the two terms. Moreover, the above-described semantic relationships are by way of example and not by way of limitation as other semantic relationships can be formed between the key term and the related terms within concept networks 206.

Figure 6 illustrates a method for the formation of concept networks 206, according to embodiments of the present invention. The description of Figure 6 is in terms of the formation of a single network within concept networks 206. This is by way

of example and not by way of limitation, as the method described in conjunction with Figure 6 relates to the generation of any number of networks within concept networks 206. Moreover, in an embodiment, each network within concept networks 206 is related to a given subject matter domain, such as computers or shoes.

5 Method 600 of Figure 6 commences with the retrieval of relevant web pages from across different web sites located on servers 104 coupled to network 102, at process block 602. In particular, network generation unit 216 traverses servers 104 across network 102 to locate the relevant web sites for the given subject matter domain for this given network within concept networks 206, as in one embodiment, each network in
10 concept networks 206 is related to a particular subject matter domain. In one embodiment, in order to determine whether a web page is considered relevant, network generation unit 216 is given certain terms, which if found in the web page would classify the web page as relevant. For example, if "computers" is the subject matter domain, these certain terms could include computer, processor, software, INTEL®, etc. However,
15 embodiments of the present invention are not so limited, as other techniques for classifying web pages as relevant can be incorporated into embodiments of the present invention. For example, a data analyst could perform a manual search across the different web sites and determine which web pages are considered relevant.

Upon retrieving the relevant web pages, network generation unit 216 identifies the
20 key terms in the relevant web pages, at process block 604. In an embodiment, prior to identifying the key terms, network generation unit 216 filters out "noise" words from these relevant web pages. "Noise" words include common articles, conjunctions and/or prepositions. Examples of common articles include "a", "an" and "the", while examples of conjunctions include "and", "or" and "nor". Moreover, examples of prepositions
25 include "of", "in" and "for". However, "noise" words are not limited to those described above, as any other words that do not carry meaning in terms of the content of the web page can be so classified.

In one embodiment, network generation unit 216 identifies these key terms employing a statistical frequency analysis of the relevant web pages. In particular, network generation unit 216 performs an analysis for each of the relevant web pages to determine the frequency of the different words and/or phrases therein. Accordingly, network generation unit 216 labels the key terms of these relevant web pages as those which occur most frequently therein. In one embodiment, the number of key terms to be identified is predetermined. For example, network generation unit 216 could identify the ten most frequent terms and/or phrases. However, embodiments of the invention are not so limited, as other techniques can be employed to determine whether a term and/or phrase occurs frequently enough to be denoted as "key." For example, in an embodiment, the key terms and/or phrases are those which occur "x" number of times within the given web page.

In one embodiment, network generation unit 216 identifies these key terms employing a weighting algorithm with regard to the relationship of two terms within a given web page. In one such embodiment, a weight related to the mutuality between two words is assigned to the terms of the given web page based on equation #1 shown below:

$$\text{(Equation \#1)} \quad MI(x,y) = f(x,y) / f(x) + f(y) - f(x,y)$$

Mutual Information (MI) for two words (x and y) define the mutuality between x and y, as is known in the art. In particular, f(x) is the occurrence frequency of word x, while f(y) is the occurrence frequency of word y. Moreover, f(x,y) is the occurrence frequency of the word pair (x,y). The higher the value of MI(x,y) is, the more likely x and y are to form a compound phrase (i.e., the key phrase) within a given web page or document.

Accordingly, network generation unit 216 identifies these key terms based on the weights related to mutuality assigned by equation #1 to the terms within the given web page. In one embodiment, network generation unit 216 employs both the statistical

frequency analysis and the weighting algorithm together to determine the key terms within each of the relevant web pages. These identified key terms are also known as anchor terms for their subsequent incorporation into the network of concept networks 206.

5 Upon identification of the key or anchor terms for a given subject matter domain, network generation unit 216 passes these terms into concordance tool 218. Accordingly, a data or text analyst uses concordance tool 218 to identify related terms to these anchor terms. In one embodiment, data analyst in conjunction with concordance tool 218 performs an analysis for a given key or anchor term separate from an analysis of other
10 such terms. In particular, concordance tool 218 retrieves lines from the relevant web pages that include the key or anchor term, at process block 606. Such lines can be sentences or portions of sentences that include the key or anchor term from the relevant web pages. Additionally, these lines are displayed to a data analyst, who can select terms related to the key or anchor term, at process block 608.

15 In particular, **Figure 7** illustrates a display of the output of concordance tool 218 used by a data analyst to identify terms related to a given anchor term. As shown, Figure 7 illustrates display 700, which includes lines 702a-702i. Lines 702a-i can include one to any number of lines across multiply display screens. While different embodiment can include different display techniques of lines 702a-702i, as shown in display 700, in one
20 embodiment, the anchor term is centered in the middle of display 700 for each of lines 702a-702i, thereby allowing the data analyst to see the context in which the anchor terms are being used.

 Additionally, in one embodiment, concordance tool 218 arranges lines 702a-702i in alphabetical order according to the word just prior to the anchor term, as illustrated in
25 display 700. In an alternative embodiment, concordance tool 218 arranges lines 702a-702i in alphabetical order according to the word just subsequent to the anchor term. These above-described orderings of lines 702a-702i are by way of example and not by

way of limitation, as embodiments of the present invention can provide other ordering techniques, such as a random order.

Upon displaying of the lines that include the anchor terms, the data analyst employing concordance tool 218 can select terms related to these anchor terms.

5 Returning to display 700, a given data analyst could select "baseball", "rubber-soled" and "spikes" from line 702a as being related terms. Moreover, in one embodiment, the data analyst defines a semantic relationship for each related term and the particular anchor term. Examples of such semantic relationships are described above in conjunction with Figure 5. Accordingly, the given data analyst selects the related terms as well as the
10 semantic relationships for a particular anchor term based on lines 702a-702i. Concordance tool 218 receives these related terms and the semantic relationships for a particular key or anchor term, at process block 610. Additionally, the above-described method 600 is reiterated for each anchor or key term within a given network within concept networks 206. Accordingly, a given network within concept networks 206 is
15 generated based on these different anchor or key terms along with their related terms and the semantic relationships there between. In particular, returning to Figure 5, network portion 500 illustrates a portion of a network for a given subject matter domain, as one to any number of key terms can be incorporated into a given network.

Returning to process block 304 of Figure 3 for the reformulation of the search
20 query, in operation, upon receipt of the initial search term or phrase, server 104 searches concept networks 206 for related terms or phrases. Returning to the example illustrated in Figure 5, if client 106 entered the initial search term "INTEL®", server 104 could revise this initial search query to include any subsidiaries, products and the stock ticker name of INTEL® stored in network portion 500. While different embodiments of the
25 present invention can couple these search terms in different formats to generate a different search query, in one embodiment, these search terms are coupled using an "AND" search operator. For example, if server 104 received the search term "INTEL®"

and located the related terms "PENTIUM®" and "INTC" in network portion 500, the reformulated search query would be "INTEL® and PENTIUM® and INTC". Accordingly for this reformulated search query, all the terms must be located in a given web page before a match is made.

5 However, embodiments of the present invention are not so limited, as any other search operator can be incorporated into the reformulated search query at process block 304. For example, other search operators to connect the initial search term with the related search terms could include OR, NOT, NEAR, etc. The search operator "OR" provides a search wherein a match is accorded when either the search term or any of the
10 related terms are located in a given web page.

 With regard to the "NOT" search operator, this operator enables a search wherein a match is accorded when a search term or related search term is specifically not found in a particular web page. Additionally, the search operator "NEAR" allows for a search wherein a match is accorded when two terms are within "x" number of words of each
15 other in the web page. Moreover, different search operators can be combined and intermixed together in the generation of a reformulated search query. For example, a combined search operator query could be: "INTEL® near PENTIUM® not INTC". Accordingly, a web page is considered a match when INTEL® is within "x" number of words of each other and when INTC is not found in the web page. The prior examples of
20 search operators and combinations thereof are by way of example and not by way of limitation, as other types of search operators as well as different combinations of search operators could be included in the reformulated search queries.

 In an embodiment, in addition to entering a search term or phrase in box 404 of web page 402, web page 402 could include a selection criteria, such as a radio box, which
25 enable client 106 to indicate the level of detail and/or number of web pages to be returned from the search query. For example, a radio box could be included in web page 402 that enables client 106 to enter one of three options for level of detail: (1) high, (2) medium or

(3) low. Accordingly, this option allows client 106 to control the number of web pages returned by the search result. In particular, if the level of detail is low, server 104 could incorporate more AND search operators into the reformulated search query in order to narrow the search. Conversely, if client 106 wants a high level of detail, server 104 could
5 incorporate more OR search operators into the reformulated search query in order to expand the search.

In one embodiment, server 104 generates the reformulated search query based on a set of parameters. For example, server 104 could be designed to incorporate all related terms into the reformulated search query using an OR search operator. In further
10 illustration, server 104 could be designed to incorporate all related items employing an AND search operator.

Moreover, in an embodiment, server 104 generates the reformulated search query based on the type of semantic relationship between the search term and the related term. In particular, web site 220 may be set up to incorporate the related terms that have certain
15 types of semantic relationships with the search term. For example, only the related terms that have a semantic relationship of "product of" may be incorporated into the reformulated search query. The above-described examples and illustration of the reformulation of the search query at process block 304 are by way of illustration and not by way of limitation, as other types methods may be employed in the reformulation of
20 this search query.

Returning to method 300, once the search query has been reformulated, database 208 is searched to locate the web pages that match the terms in the reformulated search query at process block 306. In particular, database 208 includes key data terms and phrases from particular documents from web pages across network 102. In one
25 embodiment, database 208 is generated using software programs, termed crawlers or spiders, that automatically traverse, on a periodic basis, the WWW through the different web sites thereon based on hyperlinks among the different web sites. When these

software programs locate a web site whose information is not stored in database 208 for this particular search engine, such programs transmit key data terms/phrases and other relevant information concerning the documents found on the new web site back to the web site executing these programs. Accordingly, this information about the new websites is stored into database 208 for website 220.

Upon locating the relevant documents from the web pages that have terms that match the terms in the reformulated search query, search engine 212 retrieves "N" number of such documents for display to the user of web site 220. In one embodiment, the user of web site 220 defines "N". Moreover, in an embodiment, search engine 212 orders such documents for display based on relevancy to the reformulated search query. In one such embodiment, relevancy is based on how many times the search terms and related terms in the reformulated search query are found in a given document.

Additionally, summarizer unit 214 generates a summary for each of these documents. In one embodiment, the summary is the leading text of a given document. For example, the summary could be the first 50-100 words of the particular document. In an alternative embodiment, summarizer unit 214 generates the summary based on key sentences within the given document. In one such embodiment, the weight of each sentence within a given document is calculated based on how frequent the search terms within the query are found in the sentence. Accordingly, the more frequent the search terms are found in a sentence, the more weight is given to that sentence.

Summarizer unit 214 ranks the sentences based on these calculated weights, with the higher rank being assigned to the sentences with greater weights. In one embodiment, the summary is, for example, the first five highest ranked sentences. However, embodiments of the present invention are not so limited, as other techniques can be employed to determine the summary. For example, the summary could be the first 50-100 words based on the ranking of the sentences. Accordingly, if the first two highest ranked sentences include 100 words, then these two sentences define the summary.

In an embodiment, in addition to this summary, the headlines or titles for the documents are also displayed at process block 308. In one embodiment, the key words and the related words are also displayed at process block 308. The user of search engine 212 has an opportunity to select one of the related terms, at process decision block 310.

5 If the user is satisfied with the search and the results thereof, the process is complete, at process block 314.

However, if the user wants to find different documents based on a new search, they can select one of the related terms for a new query. Search engine 212 receives this selection of a new query, at process block 312. Search engine 212 reformulates a new

10 search query based on this new term, at process block 304. Accordingly, web site 220 performs a recursive process of receiving a new search term, reformulating the query and displaying the search results until the user is satisfied with their search. Further, in an embodiment, differing weights are assigned to the search terms and the related terms in order to place different emphasis on such terms during the search. In one such

15 embodiment, search engine 212 places greater weight on those related terms that the user selected during the recursive searching process. For example, a related term, which the user selected, may only need to be found in a given document once, while the other terms in the reformulated search query may be required to be found multiple times in order for a given document to be considered a match. However, embodiments of the present

20 invention are not so limited as other techniques may be incorporated into embodiments of the present invention to provide for differing weights among the different terms of the reformulated search query.

In one embodiment, web site 220 includes conversion unit 219. Conversion unit 219 allows the user to not only read, but also hear, the search results, as conversion unit

25 219 converts the text of the search results to speech. Accordingly, any type of user, including a novice, can use embodiments of web site 220 to search across vast numbers

of web sites located on servers 104 across network 102 to find particular information in documents therein.

Moreover, the above-description was described in terms of the Internet and the WWW. However, this description was by way of example and not by way of limitation, as embodiments of the present invention are applicable to any other type of network that includes server-type computers that store data information that can be retrieved by client computers. For example, in one embodiment, network 102 could be a local area network with multiple servers and clients coupled thereto.

Memory 202 includes a machine-readable medium on which is stored a set of instructions (i.e., software) embodying any one, or all, of the methodologies described above. Software can reside, completely or at least partially, within memory 202 and/or within processing unit 204. For the purposes of this specification, the term " machine-readable medium" shall be taken to include any mechanism that provides (i.e., stores and/or transmits) information in a form readable by a machine (e.g., a computer). For example, a machine-readable medium includes read only memory (ROM); random access memory (RAM); magnetic disk storage media; optical storage media; flash memory devices; electrical, optical, acoustical or other form of propagated signals (e.g., carrier waves, infrared signals, digital signals, etc.); etc.

Although the present invention has been described with reference to specific exemplary embodiments, it will be evident that various modifications and changes may be made to these embodiments without departing from the broader spirit and scope of the invention. For example, the above-described embodiments were described in terms of the English language. However, such embodiments can be incorporated and be applicable to any other language. Accordingly, the specification and drawings are to be regarded in an illustrative rather than a restrictive sense.

What is claimed is:

1. A method comprising:
receiving a search term for a query;
searching a network of concept terms for terms related to the search term;
5 reformulating the query using the search term and the related terms;
searching a local database for data terms that match the search term and the
related terms, wherein the data terms are from documents residing on websites located on
servers across a network; and
retrieving the documents from the websites whose data terms match the search
10 term and the related terms.
2. The method of claim 1, further comprising displaying the retrieved documents,
the search terms and the related terms.
- 15 3. The method of claim 1, further comprising generating a summary of the
documents for the searched terms that match the search term and the related terms.
4. The method of claim 3, wherein the summary includes the searched terms and a
beginning portion of the documents.
- 20 5. The method of claim 1, wherein the network is the Internet.
6. The method of claim 1, wherein the network of concept terms includes links
between related terms, wherein the links are based on semantic relationships.
- 25 7. The method of claim 1, wherein the semantic relationships are selected from a
group consisting of canonical, synonym, hyponym, hypernym, part, product and member.

8. The method of claim 1, wherein related terms are more specific than the search term.

5 9. The method of claim 1, wherein different emphasis is placed on the related terms.

10. A method comprising:

recursively performing the following process until desired documents are found:

receiving a search term for a query;

10 searching a network of concept terms for terms related to the initial search term;

reformulating the query using the initial search term and the related terms;

searching a local database for data terms that match the initial search term and the related terms, wherein the data terms are from documents residing on websites

15 located on servers across a network;

displaying results of the searching of the local database; and

displaying the search term and the related terms.

11. The method of claim 10, wherein receiving the search term for the query includes
20 receiving the search term for the query based on the displaying of the search term and the related items in a prior process.

12. The method of claim 10, wherein the new search term is a related term from a prior search of the network of concept terms.

25

13. The method of claim 10, wherein reformulating the new query includes combining the new search term and the new related terms together using search operators.

5 14. The method of claim 13, wherein the search operators are selected from the group consisting of AND, OR, NOT and NEAR.

15. A method comprising:

- receiving an initial search term for a query;
- 10 searching a network of concept terms for terms related to the initial search term;
- reformulating the query using the initial search term and the related terms;
- searching a local database for data terms that match the initial search term and the related terms, wherein the data terms are from documents residing on websites located on servers across a network;
- 15 displaying results of the searching of the local database;
- displaying the search term and the related terms;
- recursively performing the following until desired documents are found:
 - receiving a new search term for a new query based on the display of the results, the search term and the related terms;
 - 20 searching the network of concept terms for new terms related to the new search term;
 - reformulating the new query using the new search term and the new related terms;
 - searching the local database for data terms that match the new search term
 - 25 and the new related terms, wherein the data terms are from documents residing on websites residing on servers across the network;
 - displaying results of the searching of the local database; and

displaying the new search term and the new related terms.

16. The method of claim 15, wherein the new search term is a related term from a prior search of the network of concept terms.

5

17. The method of claim 15, wherein reformulating the new query includes combining the new search term and the new related terms together using search operators.

10 18. The method of claim 17, wherein the search operators are selected from the group consisting of AND, OR, NOT and NEAR.

19. An apparatus comprising:
a database that includes data terms, wherein the data terms are from documents
15 residing on websites located on servers across a network;
a concept network that includes search terms and related terms that are linked together based on semantic relationships, the search terms and the related terms to locate portions of the documents based on a match-between the searchable terms and the related terms and the data terms stored in the database.

20

20. The apparatus of claim 19, wherein the semantic relationships are selected from a group consisting of canonical, synonym, hyponym, hypernym, part, product and member.

21. The apparatus of claim 19, wherein the related terms are more specific than the
25 search terms.

22. A machine-readable medium that provides instructions, which when executed by a machine, cause said machine to perform operations comprising:

- receiving a search term for a query;
- searching a network of concept terms for terms related to the search term;
- 5 reformulating the query using the search term and the related terms;
- searching a local database for data terms that match the search term and the related terms, wherein the data terms are from documents residing on websites located on servers across a network; and
- retrieving the documents from the websites whose data terms match the search
- 10 term and the related terms.

23. The machine-readable medium of claim 22, further comprising displaying the retrieved documents, the search terms and the related terms.

- 15 24. The machine-readable medium of claim 22, further comprising generating a summary of the documents for the searched terms that match the search term and the related terms.

- 25 25. The machine-readable medium of claim 22, wherein the network of concept terms includes links between related terms, wherein the links are based on semantic relationships.

- 26. The machine-readable medium of claim 22, wherein the semantic relationships are selected from a group consisting of canonical, synonym, hyponym, hypernym, part, product and member.
- 25

27. A machine-readable medium that provides instructions, which when executed by a machine, cause said machine to perform operations comprising:

recursively performing the following process until desired documents are found:

receiving a search term for a query;

5 searching a network of concept terms for terms related to the initial search term;

reformulating the query using the initial search term and the related terms;

searching a local database for data terms that match the initial search term and the related terms, wherein the data terms are from documents residing on websites

10 located on servers across a network;

displaying results of the searching of the local database; and

displaying the search term and the related terms.

28. The machine-readable medium of claim 27, wherein receiving the search term for
15 the query includes receiving the search term for the query based on the displaying of the search term and the related items in a prior process.

29. The machine-readable medium of claim 27, wherein the new search term is a related term from a prior search of the network of concept terms.

20

30. The machine-readable medium of claim 27, wherein reformulating the new query includes combining the new search term and the new related terms together using search operators.

25

ABSTRACT OF THE DISCLOSURE

According to one embodiment of the invention, a method includes receiving a search term for a query. The method also includes searching a network of concept terms for terms related to the search term. Additionally, the query is reformulated using the search term and the related terms. A local database is searched for data terms that match the search term and the related terms. The data terms are from documents residing on websites located on servers across a network. Moreover, the method includes retrieving the documents from the websites whose data terms match the search term and the related terms.

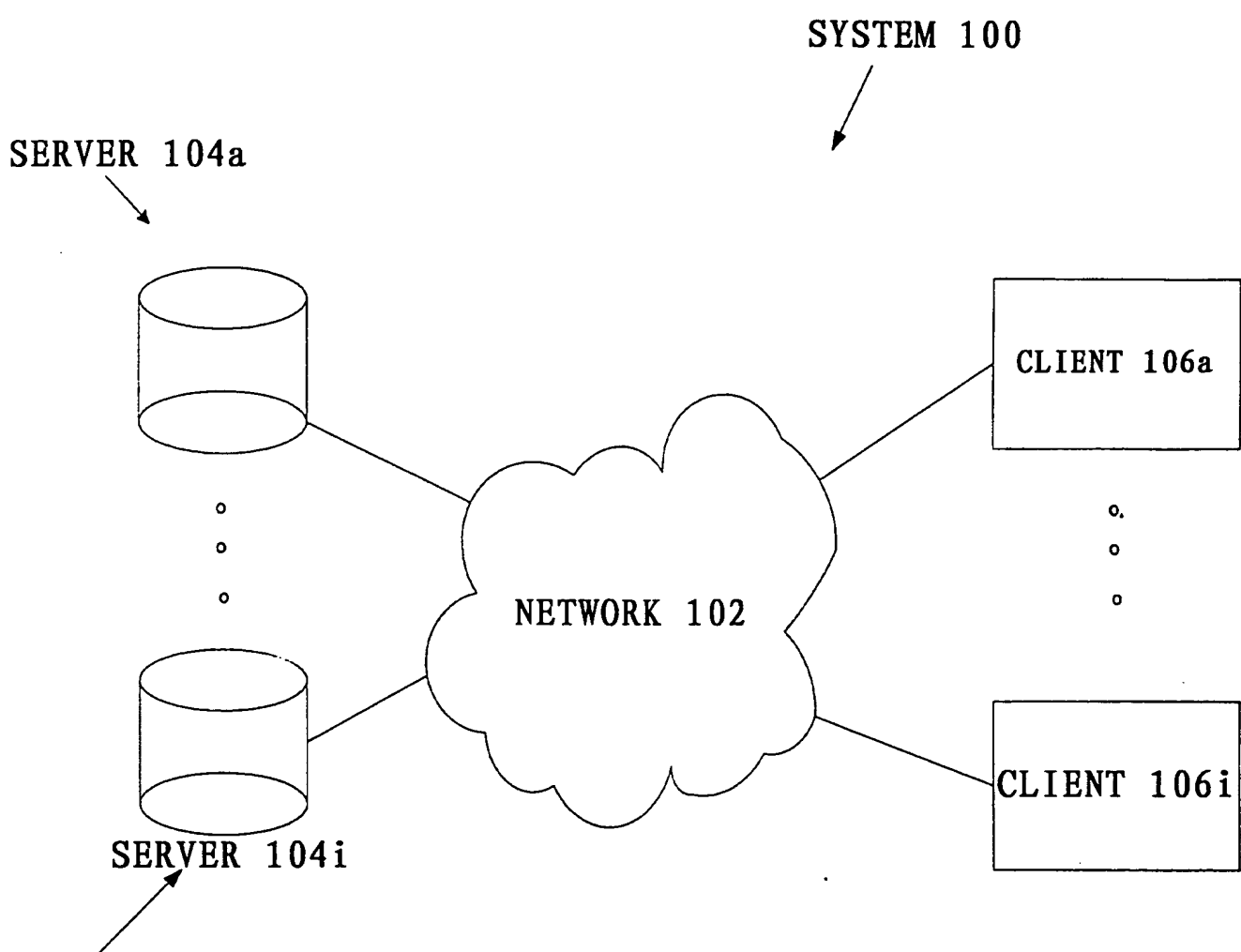


FIG. 1

SERVERS 104

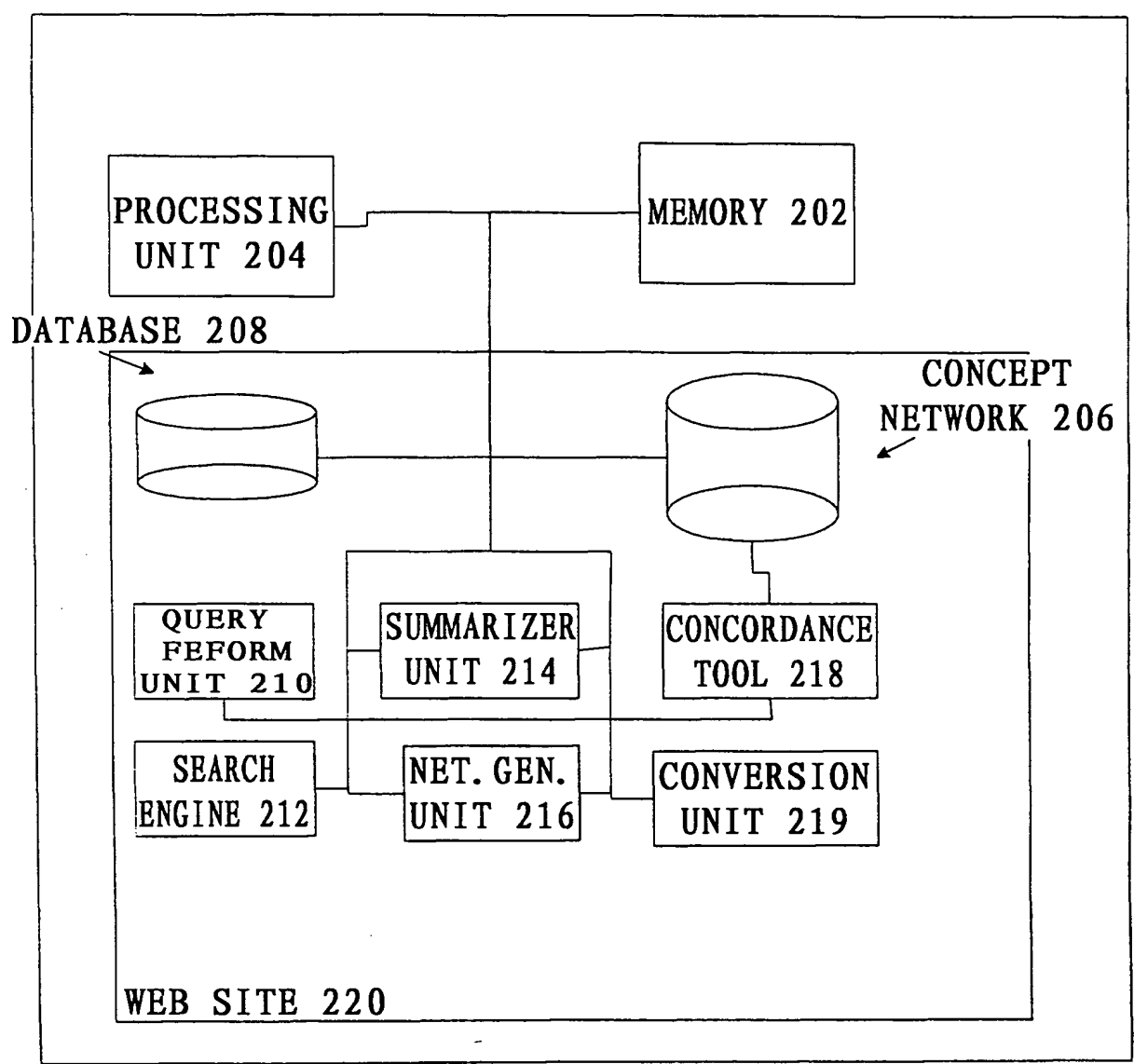
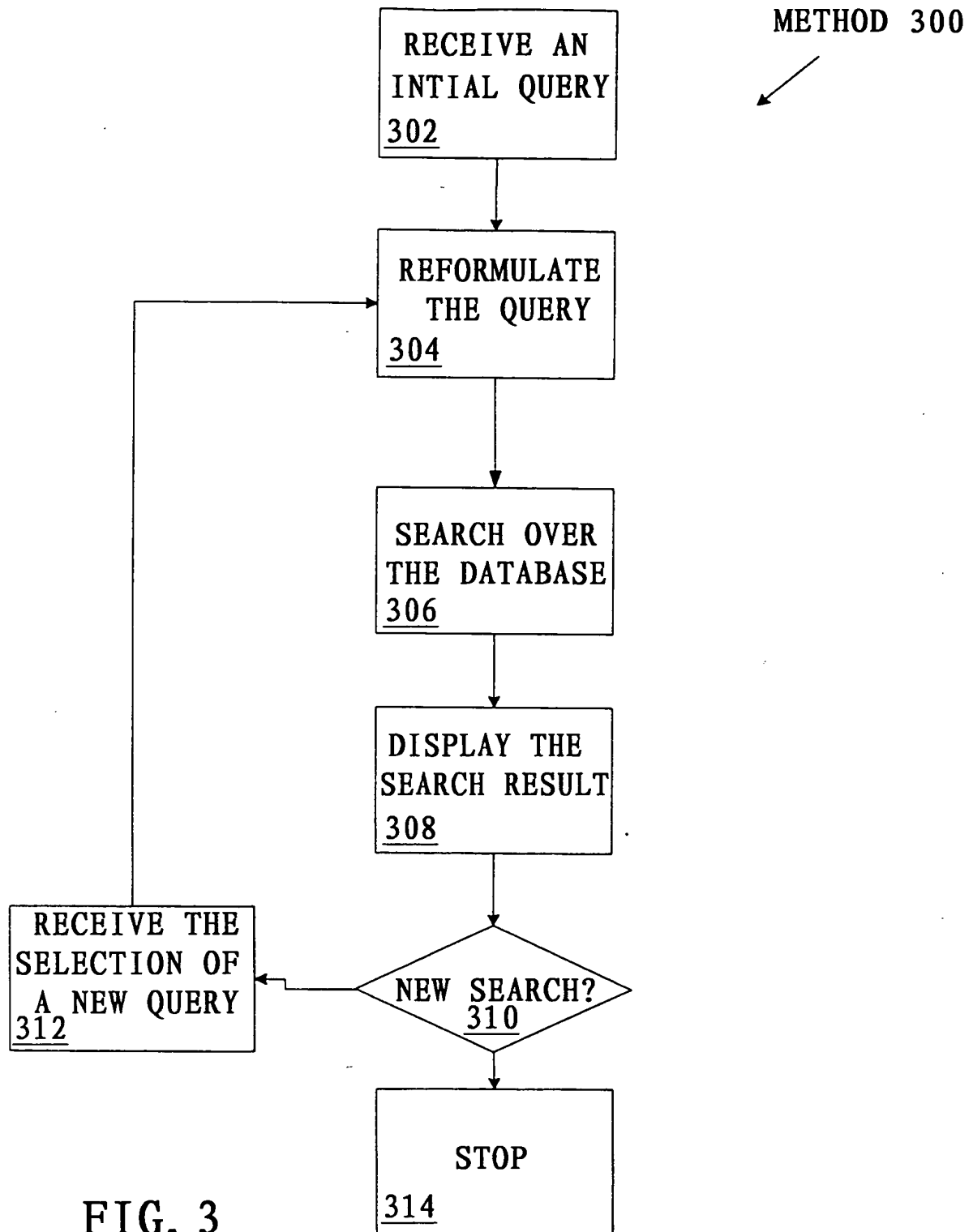


FIG. 2



4/7

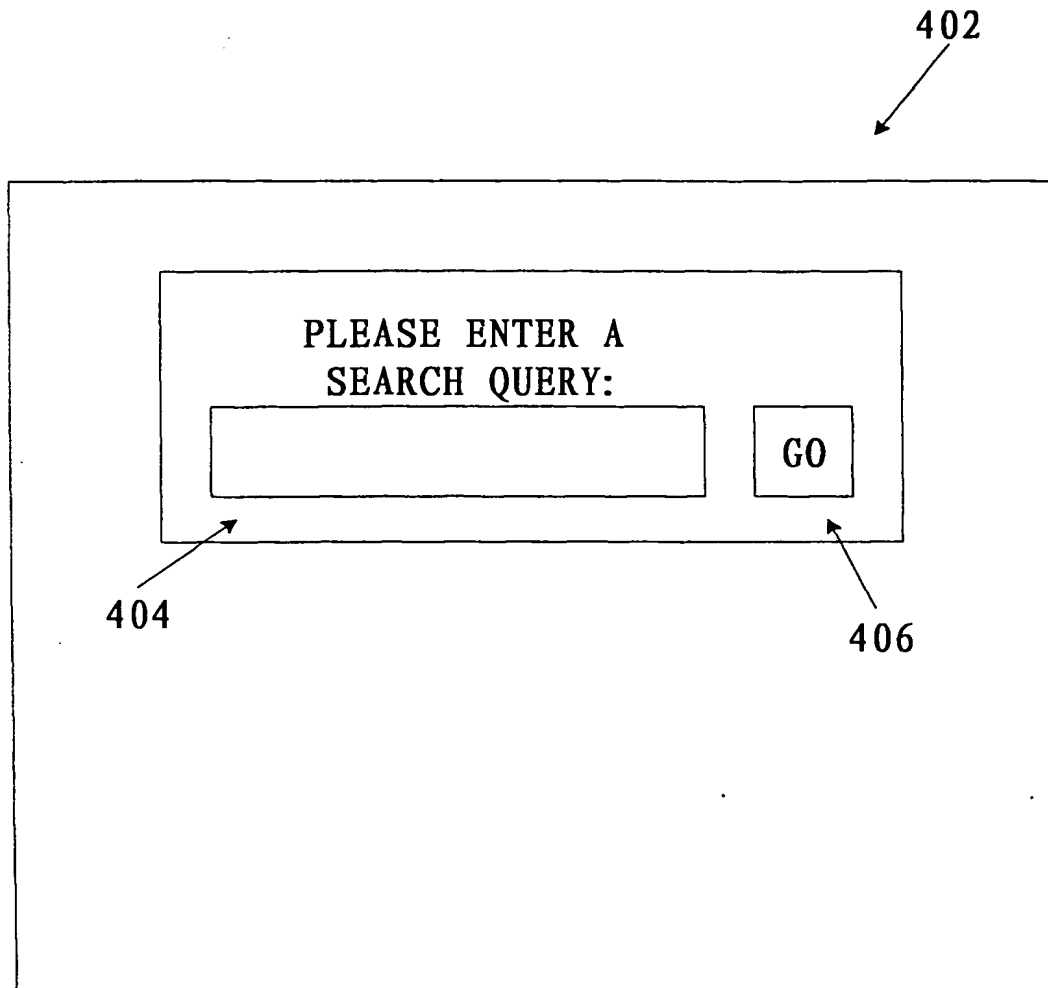


FIG. 4

替换页(细则第26条)

5/7

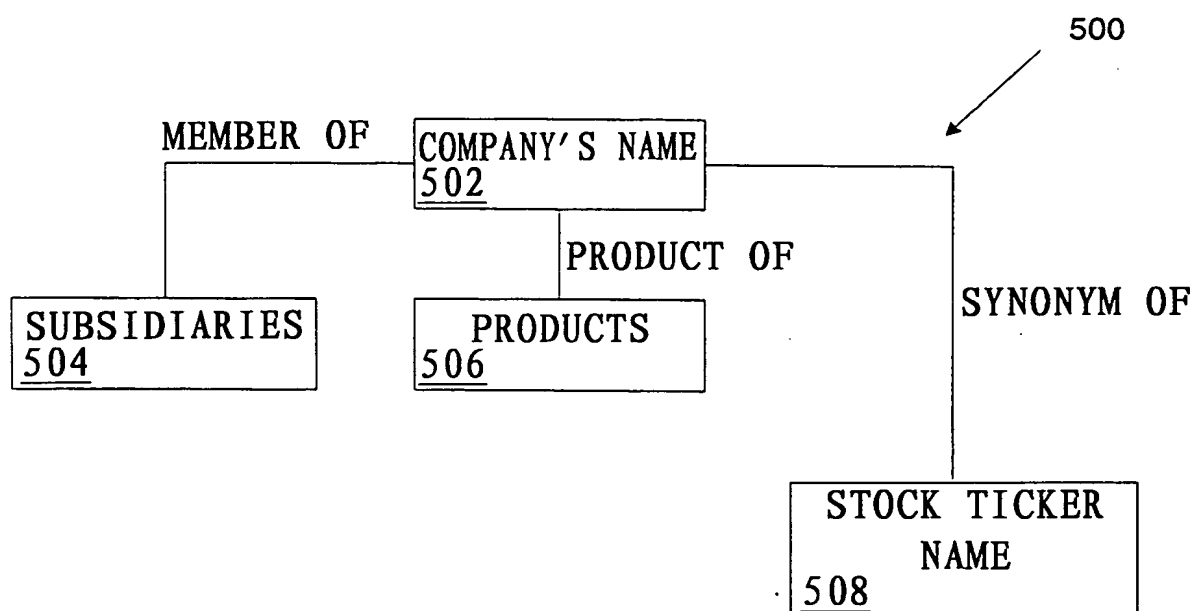


FIG. 5

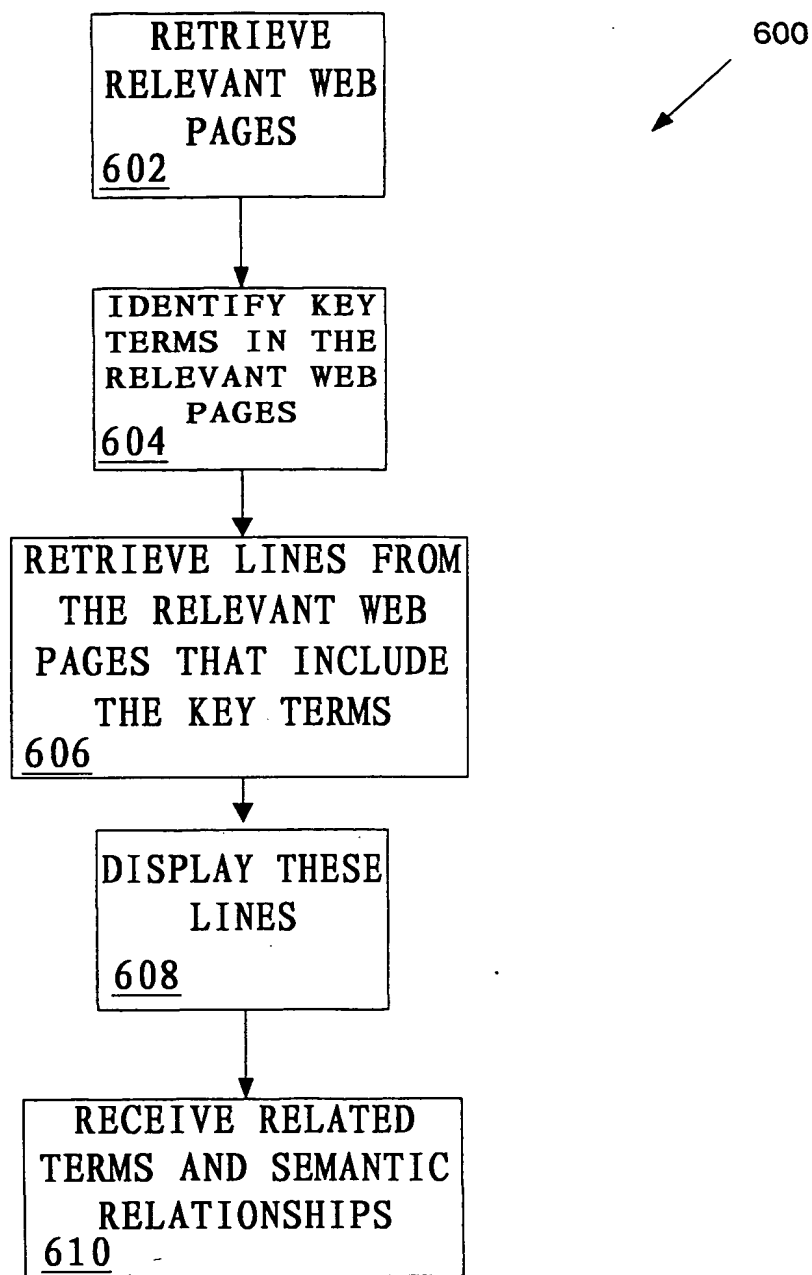


FIG. 6

7/7

DISPLAY 700

702b

702a

Baseball shoes include both rubber-soled and spikes.
The new commercial of the basketball shoes includes a superstar

o
o
o

702i

These dress shoes come in three different colors.

FIG. 7

10/018,518

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau(43) International Publication Date
4 April 2002 (04.04.2002)

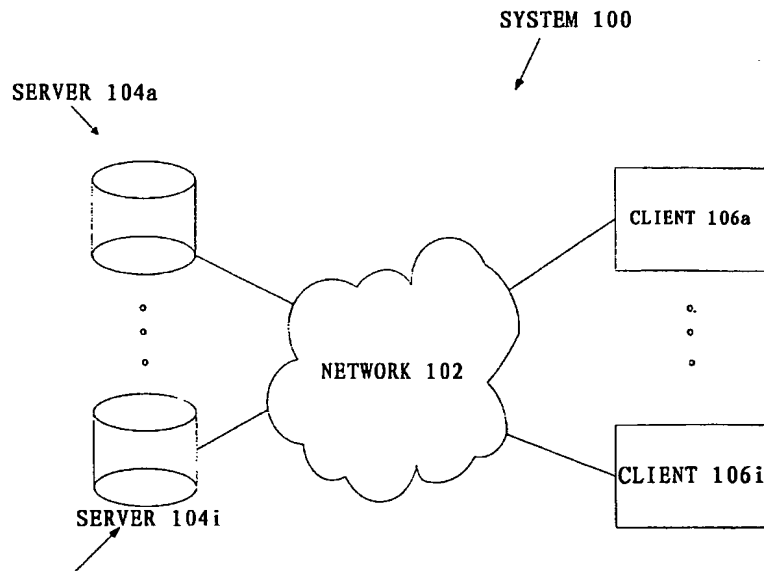
PCT

(10) International Publication Number
WO 02/27541 A1

- (51) International Patent Classification⁷: **G06F 17/30**
- (21) International Application Number: PCT/CN00/00246
- (22) International Filing Date: 23 August 2000 (23.08.2000)
- (25) Filing Language: English
- (26) Publication Language: English
- (71) Applicant (for all designated States except US): **INTEL CORPORATION** [US/US]; 2200 Mission College Boulevard, Santa Clara, CA 95052 (US).
- (72) Inventors; and
- (75) Inventors/Applicants (for US only): **ZHOU, Joe, F.** [CN/CN]; Unite 4008 Capital Mansion, No. 6 Xin Yuan Nan Road, Chaoyang District, Beijing 100004 (CN). **LIU, Weiquan** [CN/CN]; Yu zhong Xi Li 27-1204, Xicheng District, Beijing 100029 (CN).
- (74) Agent: **CCPIT PATENT AND TRADEMARK LAW OFFICE**; 8th Floor, 2 Fuchengmenwai St., Beijing 100037 (CN).
- (81) Designated States (national): AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.
- (84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).
- Published:
— with international search report

[Continued on next page]

(54) Title: A METHOD AND APPARATUS FOR CONCEPT-BASED SEARCHING ACROSS A NETWORK



(57) Abstract: According to one embodiment of the invention, a method includes receiving a search term for a query. The method also includes searching a network of concept terms for terms related to the search term. Additionally, the query is reformulated using the search term and the related terms. A local database is searched for data terms that match the search term and the related terms. The data terms are from documents residing on websites located on servers across a network. Moreover, the method includes retrieving the documents from the websites whose data terms match the search term and the related terms.

WO 02/27541 A1



For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

A METHOD AND APPARATUS FOR CONCEPT-BASED SEARCHING ACROSS A NETWORK

FIELD

5 The invention relates to searching across a network. More specifically, the invention relates to a method and apparatus for a concept-based search across a number of web sites within a network.

BACKGROUND

10 With the advent of the Internet and the World Wide Web (WWW), the vast amount of information available across the different web sites on the Internet continues to grow as the number of networks and servers connected thereto continue to increase on a global scale. Accordingly, the availability of this information on the Internet enables users to obtain such information quickly. However, locating such information on the
15 Internet can sometimes be difficult because of the vast numbers of websites available.

 Currently, particular web sites on the WWW include search mechanisms termed "search engines" that have been developed in order to assist the users of the WWW in locating specific information therein. In particular, software programs, termed crawlers or spiders, automatically traverse, on a periodic basis, the WWW through the different
20 web sites located thereon based on hyperlinks among the different web sites. When these software programs locate a web site whose information is not stored in the local database for this particular search engine, such programs transmit keywords and other relevant information concerning the new web site back to the web site executing these programs. Accordingly, this information about the new websites is stored into the local database for
25 the search engine website.

 Moreover, because of the dynamic nature of the WWW, these software programs also update such relevant information concerning a web site whose information is already

stored in the search engine's local database. Accordingly, these search engine web sites store these search terms in a local database with other search terms from other websites. In operation, a user of these search engines enters a search query into the search engine and based on this local database, the search engine locates web pages on different web sites whose search terms include the search query and retrieves portions of these web pages. The search engine then displays these portions of the web pages to the user, thereby allowing the user to select from among the different web pages.

Additionally, another type of search engine for the WWW performs a search among the search engine results described above. In particular, these type of search engines perform a search among the individual search engines as described above and combines the results from these individual search engines into a combined results, thereby allowing the user to obtain results from across multiple search engines for a single search query.

15 BRIEF DESCRIPTION OF THE DRAWINGS

Embodiments of the invention may be best understood by referring to the following description and accompanying drawings which illustrate such embodiments. In the drawings:

Figure 1 is block diagram illustrating a system that incorporates embodiments of the present invention;

Figure 2 illustrates one of servers 104*a-i* that include embodiments of the present invention;

Figure 3 is a flowchart illustrating a method for performing an expansive search for data information across servers 104 coupled to network 102, according to embodiments of the present invention;

Figure 4 illustrates a web page for entering a search term or phrase by clients 106;

Figure 5 illustrates an example of a portion of a network of conceptual terms, according to embodiments of the present invention;

Figure 6 illustrates a method for the formation of concept networks 206, according to embodiments of the present invention; and

5 **Figure 7** illustrates a display of the output of concordance tool 218 used by a data analyst to identify terms related to a given anchor term.

Detailed Description

10 In the following description, numerous specific details are set forth to provide a thorough understanding of the invention. However, it is understood that the invention may be practiced without these specific details. In other instances, well-known structures and techniques have not been shown in detail in order not to obscure embodiments of the present invention.

15 **Figure 1** is block diagram illustrating a system that incorporates embodiments of the present invention. In particular, Figure 1 illustrates system 100 that includes servers 104*a-i* and clients 106*a-i*, which are coupled together through network 102. Servers 104*a-i* can include one to any number of such servers in system 100. In an embodiment, servers 104*a-i* are host computers, which store different types of data information
20 including one to a number of different web sites with each web site have one to a number of web pages thereon. Additionally, clients 106*a-i* can include one to any number of such clients in system 100. Clients 106*a-i* are client computers that have the network capability to connect to servers 104*a-i* through network 102.

25 While different embodiments of the present invention could include different types of networks, in an embodiment, network 102 is the Internet that includes the WWW. However, embodiments of the present invention could include other types of networks. For example, in one embodiment, network 102 is a local area network (LAN).

In another embodiment, network 106 is a wide area network (WAN). Further, network 106 can be a combination of different networks that provide communication among servers 104*a-i* and clients 106*a-i*. In operation for a given client, such as client 106*a*, connecting to a given server, such as server 104*a*, client 106*a* connects to server 104*a* through network 102 and is able to download and upload different web pages from and to different websites residing on server 104*a*.

Figure 2 illustrates one of servers 104*a-i* (hereinafter "servers 104") that include embodiments of the present invention. As shown, servers 104 include processing unit 204, memory 202 and website 220. Additionally, website 220 includes database 208, concept networks 206, query reformulation unit 210, summarizer unit 212, search engine 214, network generation unit 216 and concordance tool 218, which are coupled together as shown. The following description of servers 104 is in terms of a particular web site on servers 104. However, the capability and/or functionality of servers 104 are not so limited, as servers 104 can service and/or host a number of web sites, which clients 106*a-i* (hereinafter "clients 106") can access across network 102.

Database 208 and concept networks 206 are associated with website 220, which is stored on servers 104. In particular, website 220 acts as a search engine that enables clients 106 to search web pages on the local server, which stores this website, as well as remote servers coupled to network 102. In particular, website 220 is a search engine that allows the users thereof to enter search terms on a given web page and have this search engine return the results of a search. Additionally, concept networks 206 are defined to include one to any number of networks, whose creation is described in more detail below in conjunction with Figure 6. In an embodiment, each network within concept networks 206 is related to a given subject matter domain. For example, one network relates to computers, while another network in concept networks 206 relates to shoes.

Moreover, servers 104 include query reformation unit 210, search engine unit 212, summarizer 214, network generation unit 216 and concordance tool 218, which, in

one embodiment, are software programs that reside in memory 202 and processing unit 204 during their execution on processing unit 204. Query reformation unit 210, search engine unit 212, summarizer 214, network generation unit 216 and concordance tool 218 will be described in more detail below. In conjunction with system 100 of Figure 1 and
5 the more detailed illustration of servers 104 of Figure 2, a method of operation of system 100 in accordance with embodiments of the present invention will now be described with reference to the flow chart shown in Figure 3.

In particular, **Figure 3** is a flowchart illustrating a method for performing an expansive search for data information, such as documents within web pages, across
10 servers 104 coupled to network 102, according to embodiments of the present invention. Figure 3 illustrates method 300 that commences with the receipt of an initial search query, at process block 302. **Figure 4** illustrates a web page for entering a search term or phrase by clients 106. In particular, Figure 4 illustrates web page 402 that includes box 404 for entering the search term or phrase and go button 406 to initialize the search.
15 Once client 106 enters a search term or phrase into box 404 and presses go button 406, web page 402 is transmitted back to server 104, which receives this initial search query, as shown at process block 302.

Upon receiving this initial search query, server 104 reformulates this query, at process block 304. In particular, server 104 searches a network within concept networks
20 206 to find terms related to the initial search term or phrase. **Figure 5** illustrates an example of a portion of a network within concept networks 206, according to embodiments of the present invention. As shown, Figure 5 includes network portion 500 that includes company's name 502, the company's subsidiaries 504, the company's products 506 and the company's stock ticker name 508, which are linked together. For
25 example, network portion 500 could include the term "INTEL[®]" (company's name 502). Accordingly, company's products 506 could include the term "PENTIUM[®]" as one of the

name of the processors that INTEL® produces. Moreover, network portion 500, in this example, would include "INTC" as stock ticker name 508.

Moreover as shown in Figure 5, each pair of related terms in the network is connected by a link that specifies the semantic relationship between the related terms. As illustrated, products 506 are products of company's name 502. Additionally, subsidiaries 504 are members of company's name 502. Network portion 500 also includes a synonym relationship for denoting a similar name or phrase to a given name. In particular, stock ticker name 508 is a synonym of company's name 502. Embodiments of the present invention are not limited to the examples of semantic relationships illustrated in network portion 500. For example, another type of semantic relationship that can be incorporated into embodiments of the present invention is a canonical one. In particular, a canonical relationship is defined as the simplest form of a given word. For example, the simplest form of a telephone is a communication device.

"Hyponym of" is another example of a semantic relationship. A hyponym relationship denotes a relationship of a subcategory to a more general class. An illustration of a hyponym is a chair being a subcategory of furniture. Moreover, another semantic relationship is a hypernym one, which is a relationship of a general class to a subcategory. For example, automobile is a hypernym of a car. Further, another semantic relationship between the terms within the network is a "part" relationship. For example a bicycle tire is part of a bicycle. Accordingly, each link with concept networks 206 includes a description of the semantic relationship between the two terms. Moreover, the above-described semantic relationships are by way of example and not by way of limitation as other semantic relationships can be formed between the key term and the related terms within concept networks 206.

Figure 6 illustrates a method for the formation of concept networks 206, according to embodiments of the present invention. The description of Figure 6 is in terms of the formation of a single network within concept networks 206. This is by way

of example and not by way of limitation, as the method described in conjunction with Figure 6 relates to the generation of any number of networks within concept networks 206. Moreover, in an embodiment, each network within concept networks 206 is related to a given subject matter domain, such as computers or shoes.

5 Method 600 of Figure 6 commences with the retrieval of relevant web pages from across different web sites located on servers 104 coupled to network 102, at process block 602. In particular, network generation unit 216 traverses servers 104 across network 102 to locate the relevant web sites for the given subject matter domain for this given network within concept networks 206, as in one embodiment, each network in
10 concept networks 206 is related to a particular subject matter domain. In one embodiment, in order to determine whether a web page is considered relevant, network generation unit 216 is given certain terms, which if found in the web page would classify the web page as relevant. For example, if "computers" is the subject matter domain, these certain terms could include computer, processor, software, INTEL[®], etc. However,
15 embodiments of the present invention are not so limited, as other techniques for classifying web pages as relevant can be incorporated into embodiments of the present invention. For example, a data analyst could perform a manual search across the different web sites and determine which web pages are considered relevant.

 Upon retrieving the relevant web pages, network generation unit 216 identifies the
20 key terms in the relevant web pages, at process block 604. In an embodiment, prior to identifying the key terms, network generation unit 216 filters out "noise" words from these relevant web pages. "Noise" words include common articles, conjunctions and/or prepositions. Examples of common articles include "a", "an" and "the", while examples of conjunctions include "and", "or" and "nor". Moreover, examples of prepositions
25 include "of", "in" and "for". However, "noise" words are not limited to those described above, as any other words that do not carry meaning in terms of the content of the web page can be so classified.

In one embodiment, network generation unit 216 identifies these key terms employing a statistical frequency analysis of the relevant web pages. In particular, network generation unit 216 performs an analysis for each of the relevant web pages to determine the frequency of the different words and/or phrases therein. Accordingly, network generation unit 216 labels the key terms of these relevant web pages as those which occur most frequently therein. In one embodiment, the number of key terms to be identified is predetermined. For example, network generation unit 216 could identify the ten most frequent terms and/or phrases. However, embodiments of the invention are not so limited, as other techniques can be employed to determine whether a term and/or phrase occurs frequently enough to be denoted as "key." For example, in an embodiment, the key terms and/or phrases are those which occur "x" number of times within the given web page.

In one embodiment, network generation unit 216 identifies these key terms employing a weighting algorithm with regard to the relationship of two terms within a given web page. In one such embodiment, a weight related to the mutuality between two words is assigned to the terms of the given web page based on equation #1 shown below:

$$\text{(Equation \#1)} \quad MI(x,y) = f(x,y) / f(x) + f(y) - f(x,y)$$

Mutual Information (MI) for two words (x and y) define the mutuality between x and y, as is known in the art. In particular, f(x) is the occurrence frequency of word x, while f(y) is the occurrence frequency of word y. Moreover, f(x,y) is the occurrence frequency of the word pair (x,y). The higher the value of MI(x,y) is, the more likely x and y are to form a compound phrase (i.e., the key phrase) within a given web page or document.

Accordingly, network generation unit 216 identifies these key terms based on the weights related to mutuality assigned by equation #1 to the terms within the given web page. In one embodiment, network generation unit 216 employs both the statistical

frequency analysis and the weighting algorithm together to determine the key terms within each of the relevant web pages. These identified key terms are also known as anchor terms for their subsequent incorporation into the network of concept networks 206.

5 Upon identification of the key or anchor terms for a given subject matter domain, network generation unit 216 passes these terms into concordance tool 218. Accordingly, a data or text analyst uses concordance tool 218 to identify related terms to these anchor terms. In one embodiment, data analyst in conjunction with concordance tool 218 performs an analysis for a given key or anchor term separate from an analysis of other
10 such terms. In particular, concordance tool 218 retrieves lines from the relevant web pages that include the key or anchor term, at process block 606. Such lines can be sentences or portions of sentences that include the key or anchor term from the relevant web pages. Additionally, these lines are displayed to a data analyst, who can select terms related to the key or anchor term, at process block 608.

15 In particular, **Figure 7** illustrates a display of the output of concordance tool 218 used by a data analyst to identify terms related to a given anchor term. As shown, Figure 7 illustrates display 700, which includes lines 702a-702i. Lines 702a-i can include one to any number of lines across multiply display screens. While different embodiment can include different display techniques of lines 702a-702i, as shown in display 700, in one
20 embodiment, the anchor term is centered in the middle of display 700 for each of lines 702a-702i, thereby allowing the data analyst to see the context in which the anchor terms are being used.

 Additionally, in one embodiment, concordance tool 218 arranges lines 702a-702i in alphabetical order according to the word just prior to the anchor term, as illustrated in
25 display 700. In an alternative embodiment, concordance tool 218 arranges lines 702a-702i in alphabetical order according to the word just subsequent to the anchor term. These above-described orderings of lines 702a-702i are by way of example and not by

way of limitation, as embodiments of the present invention can provide other ordering techniques, such as a random order.

Upon displaying of the lines that include the anchor terms, the data analyst employing concordance tool 218 can select terms related to these anchor terms.

5 Returning to display 700, a given data analyst could select "baseball", "rubber-soled" and "spikes" from line 702a as being related terms. Moreover, in one embodiment, the data analyst defines a semantic relationship for each related term and the particular anchor term. Examples of such semantic relationships are described above in conjunction with Figure 5. Accordingly, the given data analyst selects the related terms as well as the
10 semantic relationships for a particular anchor term based on lines 702a-702i.

Concordance tool 218 receives these related terms and the semantic relationships for a particular key or anchor term, at process block 610. Additionally, the above-described method 600 is reiterated for each anchor or key term within a given network within
— concept networks 206. Accordingly, a given network within concept networks 206 is
15 generated based on these different anchor or key terms along with their related terms and the semantic relationships there between. In particular, returning to Figure 5, network portion 500 illustrates a portion of a network for a given subject matter domain, as one to any number of key terms can be incorporated into a given network.

Returning to process block 304 of Figure 3 for the reformulation of the search
20 query, in operation, upon receipt of the initial search term or phrase, server 104 searches concept networks 206 for related terms or phrases. Returning to the example illustrated in Figure 5, if client 106 entered the initial search term "INTEL[®]", server 104 could revise this initial search query to include any subsidiaries, products and the stock ticker name of INTEL[®] stored in network portion 500. While different embodiments of the
25 present invention can couple these search terms in different formats to generate a different search query, in one embodiment, these search terms are coupled using an "AND" search operator. For example, if server 104 received the search term "INTEL[®]"

and located the related terms "PENTIUM[®]" and "INTC" in network portion 500, the reformulated search query would be "INTEL[®] and PENTIUM[®] and INTC". Accordingly for this reformulated search query, all the terms must be located in a given web page before a match is made.

5 However, embodiments of the present invention are not so limited, as any other search operator can be incorporated into the reformulated search query at process block 304. For example, other search operators to connect the initial search term with the related search terms could include OR, NOT, NEAR, etc. The search operator "OR" provides a search wherein a match is accorded when either the search term or any of the
10 related terms are located in a given web page.

 With regard to the "NOT" search operator, this operator enables a search wherein a match is accorded when a search term or related search term is specifically not found in a particular web page. Additionally, the search operator "NEAR" allows for a search wherein a match is accorded when two terms are within "x" number of words of each
15 other in the web page. Moreover, different search operators can be combined and intermixed together in the generation of a reformulated search query. For example, a combined search operator query could be: "INTEL[®] near PENTIUM[®] not INTC". Accordingly, a web page is considered a match when INTEL[®] is within "x" number of words of each other and when INTC is not found in the web page. The prior examples of
20 search operators and combinations thereof are by way of example and not by way of limitation, as other types of search operators as well as different combinations of search operators could be included in the reformulated search queries.

 In an embodiment, in addition to entering a search term or phrase in box 404 of web page 402, web page 402 could include a selection criteria, such as a radio box, which
25 enable client 106 to indicate the level of detail and/or number of web pages to be returned from the search query. For example, a radio box could be included in web page 402 that enables client 106 to enter one of three options for level of detail: (1) high, (2) medium or

(3) low. Accordingly, this option allows client 106 to control the number of web pages returned by the search result. In particular, if the level of detail is low, server 104 could incorporate more AND search operators into the reformulated search query in order to narrow the search. Conversely, if client 106 wants a high level of detail, server 104 could incorporate more OR search operators into the reformulated search query in order to expand the search.

In one embodiment, server 104 generates the reformulated search query based on a set of parameters. For example, server 104 could be designed to incorporate all related terms into the reformulated search query using an OR search operator. In further illustration, server 104 could be designed to incorporate all related items employing an AND search operator.

Moreover, in an embodiment, server 104 generates the reformulated search query based on the type of semantic relationship between the search term and the related term. In particular, web site 220 may be set up to incorporate the related terms that have certain types of semantic relationships with the search term. For example, only the related terms that have a semantic relationship of "product of" may be incorporated into the reformulated search query. The above-described examples and illustration of the reformulation of the search query at process block 304 are by way of illustration and not by way of limitation, as other types methods may be employed in the reformulation of this search query.

Returning to method 300, once the search query has been reformulated, database 208 is searched to locate the web pages that match the terms in the reformulated search query at process block 306. In particular, database 208 includes key data terms and phrases from particular documents from web pages across network 102. In one embodiment, database 208 is generated using software programs, termed crawlers or spiders, that automatically traverse, on a periodic basis, the WWW through the different web sites thereon based on hyperlinks among the different web sites. When these

software programs locate a web site whose information is not stored in database 208 for this particular search engine, such programs transmit key data terms/phrases and other relevant information concerning the documents found on the new web site back to the web site executing these programs. Accordingly, this information about the new websites
5 is stored into database 208 for website 220.

Upon locating the relevant documents from the web pages that have terms that match the terms in the reformulated search query, search engine 212 retrieves "N" number of such documents for display to the user of web site 220. In one embodiment, the user of web site 220 defines "N". Moreover, in an embodiment, search engine 212
10 orders such documents for display based on relevancy to the reformulated search query. In one such embodiment, relevancy is based on how many times the search terms and related terms in the reformulated search query are found in a given document.

Additionally, summarizer unit 214 generates a summary for each of these documents. In one embodiment, the summary is the leading text of a given document.
15 For example, the summary could be the first 50-100 words of the particular document. In an alternative embodiment, summarizer unit 214 generates the summary based on key sentences within the given document. In one such embodiment, the weight of each sentence within a given document is calculated based on how frequent the search terms within the query are found in the sentence. Accordingly, the more frequent the search
20 terms are found in a sentence, the more weight is given to that sentence.

Summarizer unit 214 ranks the sentences based on these calculated weights, with the higher rank being assigned to the sentences with greater weights. In one embodiment, the summary is, for example, the first five highest ranked sentences. However, —
embodiments of the present invention are not so limited, as other techniques can be
25 employed to determine the summary. For example, the summary could be the first 50-100 words based on the ranking of the sentences. Accordingly, if the first two highest ranked sentences include 100 words, then these two sentences define the summary.

In an embodiment, in addition to this summary, the headlines or titles for the documents are also displayed at process block 308. In one embodiment, the key words and the related words are also displayed at process block 308. The user of search engine 212 has an opportunity to select one of the related terms, at process decision block 310.

5 If the user is satisfied with the search and the results thereof, the process is complete, at process block 314.

However, if the user wants to find different documents based on a new search, they can select one of the related terms for a new query. Search engine 212 receives this selection of a new query, at process block 312. Search engine 212 reformulates a new
10 search query based on this new term, at process block 304. Accordingly, web site 220 performs a recursive process of receiving a new search term, reformulating the query and displaying the search results until the user is satisfied with their search. Further, in an embodiment, differing weights are assigned to the search terms and the related terms in order to place different emphasis on such terms during the search. In one such
15 embodiment, search engine 212 places greater weight on those related terms that the user selected during the recursive searching process. For example, a related term, which the user selected, may only need to be found in a given document once, while the other terms in the reformulated search query may be required to be found multiple times in order for a given document to be considered a match. However, embodiments of the present
20 invention are not so limited as other techniques may be incorporated into embodiments of the present invention to provide for differing weights among the different terms of the reformulated search query.

In one embodiment, web site 220 includes conversion unit 219. Conversion unit 219 allows the user to not only read, but also hear, the search results, as conversion unit
25 219 converts the text of the search results to speech. Accordingly, any type of user, including a novice, can use embodiments of web site 220 to search across vast numbers

of web sites located on servers 104 across network 102 to find particular information in documents therein.

Moreover, the above-description was described in terms of the Internet and the WWW. However, this description was by way of example and not by way of limitation, as embodiments of the present invention are applicable to any other type of network that includes server-type computers that store data information that can be retrieved by client computers. For example, in one embodiment, network 102 could be a local area network with multiple servers and clients coupled thereto.

Memory 202 includes a machine-readable medium on which is stored a set of instructions (i.e., software) embodying any one, or all, of the methodologies described above. Software can reside, completely or at least partially, within memory 202 and/or within processing unit 204. For the purposes of this specification, the term " machine-readable medium" shall be taken to include any mechanism that provides (i.e., stores and/or transmits) information in a form readable by a machine (e.g., a computer). For example, a machine-readable medium includes read only memory (ROM); random access memory (RAM); magnetic disk storage media; optical storage media; flash memory devices; electrical, optical, acoustical or other form of propagated signals (e.g., carrier waves, infrared signals, digital signals, etc.); etc.

Although the present invention has been described with reference to specific exemplary embodiments, it will be evident that various modifications and changes may be made to these embodiments without departing from the broader spirit and scope of the invention. For example, the above-described embodiments were described in terms of the English language. However, such embodiments can be incorporated and be applicable to any other language. Accordingly, the specification and drawings are to be regarded in an illustrative rather than a restrictive sense.

What is claimed is:

1. A method comprising:
receiving a search term for a query;
searching a network of concept terms for terms related to the search term;
5 reformulating the query using the search term and the related terms;
searching a local database for data terms that match the search term and the
related terms, wherein the data terms are from documents residing on websites located on
servers across a network; and
retrieving the documents from the websites whose data terms match the search
10 term and the related terms.
2. The method of claim 1, further comprising displaying the retrieved documents,
the search terms and the related terms.
- 15 3. The method of claim 1, further comprising generating a summary of the
documents for the searched terms that match the search term and the related terms.
4. The method of claim 3, wherein the summary includes the searched terms and a
beginning portion of the documents.
20
5. The method of claim 1, wherein the network is the Internet.
6. The method of claim 1, wherein the network of concept terms includes links
between related terms, wherein the links are based on semantic relationships.
25
7. The method of claim 1, wherein the semantic relationships are selected from a
group consisting of canonical, synonym, hyponym, hypernym, part, product and member.

8. The method of claim 1, wherein related terms are more specific than the search term.

5 9. The method of claim 1, wherein different emphasis is placed on the related terms.

10. A method comprising:

recursively performing the following process until desired documents are found:

receiving a search term for a query;

10 searching a network of concept terms for terms related to the initial search term;

reformulating the query using the initial search term and the related terms;

searching a local database for data terms that match the initial search term

and the related terms, wherein the data terms are from documents residing on websites

15 located on servers across a network;

displaying results of the searching of the local database; and

displaying the search term and the related terms.

11. The method of claim 10, wherein receiving the search term for the query includes
20 receiving the search term for the query based on the displaying of the search term and the related items in a prior process.

12. The method of claim 10, wherein the new search term is a related term from a
prior search of the network of concept terms.

25

13. The method of claim 10, wherein reformulating the new query includes combining the new search term and the new related terms together using search operators.

5 14. The method of claim 13, wherein the search operators are selected from the group consisting of AND, OR, NOT and NEAR.

15. A method comprising:
receiving an initial search term for a query;
10 searching a network of concept terms for terms related to the initial search term;
reformulating the query using the initial search term and the related terms;
searching a local database for data terms that match the initial search term and the related terms, wherein the data terms are from documents residing on websites located on servers across a network;
15 displaying results of the searching of the local database;
displaying the search term and the related terms;
recursively performing the following until desired documents are found:
receiving a new search term for a new query based on the display of the results, the search term and the related terms;
20 searching the network of concept terms for new terms related to the new search term;
reformulating the new query using the new search term and the new related terms;
searching the local database for data terms that match the new search term
25 and the new related terms, wherein the data terms are from documents residing on websites residing on servers across the network;
displaying results of the searching of the local database; and

displaying the new search term and the new related terms.

16. The method of claim 15, wherein the new search term is a related term from a prior search of the network of concept terms.

5

17. The method of claim 15, wherein reformulating the new query includes combining the new search term and the new related terms together using search operators.

10 18. The method of claim 17, wherein the search operators are selected from the group consisting of AND, OR, NOT and NEAR.

19. An apparatus comprising:

15 a database that includes data terms, wherein the data terms are from documents residing on websites located on servers across a network;

a concept network that includes search terms and related terms that are linked together based on semantic relationships, the search terms and the related terms to locate portions of the documents based on a match between the searchable terms and the related terms and the data terms stored in the database.

20

20. The apparatus of claim 19, wherein the semantic relationships are selected from a group consisting of canonical, synonym, hyponym, hypernym, part, product and member.

21. The apparatus of claim 19, wherein the related terms are more specific than the
25 search terms.

22. A machine-readable medium that provides instructions, which when executed by a machine, cause said machine to perform operations comprising:

receiving a search term for a query;

searching a network of concept terms for terms related to the search term;

5 reformulating the query using the search term and the related terms;

searching a local database for data terms that match the search term and the related terms, wherein the data terms are from documents residing on websites located on servers across a network; and

10 retrieving the documents from the websites whose data terms match the search term and the related terms.

23. The machine-readable medium of claim 22, further comprising displaying the retrieved documents, the search terms and the related terms.

15 24. The machine-readable medium of claim 22, further comprising generating a summary of the documents for the searched terms that match the search term and the related terms.

25. The machine-readable medium of claim 22, wherein the network of concept terms includes links between related terms, wherein the links are based on semantic relationships.

26. The machine-readable medium of claim 22, wherein the semantic relationships are selected from a group consisting of canonical, synonym, hyponym, hypernym, part, product and member.

27. A machine-readable medium that provides instructions, which when executed by a machine, cause said machine to perform operations comprising:

recursively performing the following process until desired documents are found:

receiving a search term for a query;

5 searching a network of concept terms for terms related to the initial search term;

reformulating the query using the initial search term and the related terms;

searching a local database for data terms that match the initial search term

and the related terms, wherein the data terms are from documents residing on websites

10 located on servers across a network;

displaying results of the searching of the local database; and

displaying the search term and the related terms.

28. The machine-readable medium of claim 27, wherein receiving the search term for
15 the query includes receiving the search term for the query based on the displaying of the search term and the related items in a prior process.

29. The machine-readable medium of claim 27, wherein the new search term is a related term from a prior search of the network of concept terms.

20

30. The machine-readable medium of claim 27, wherein reformulating the new query includes combining the new search term and the new related terms together using search operators.

25

1/7

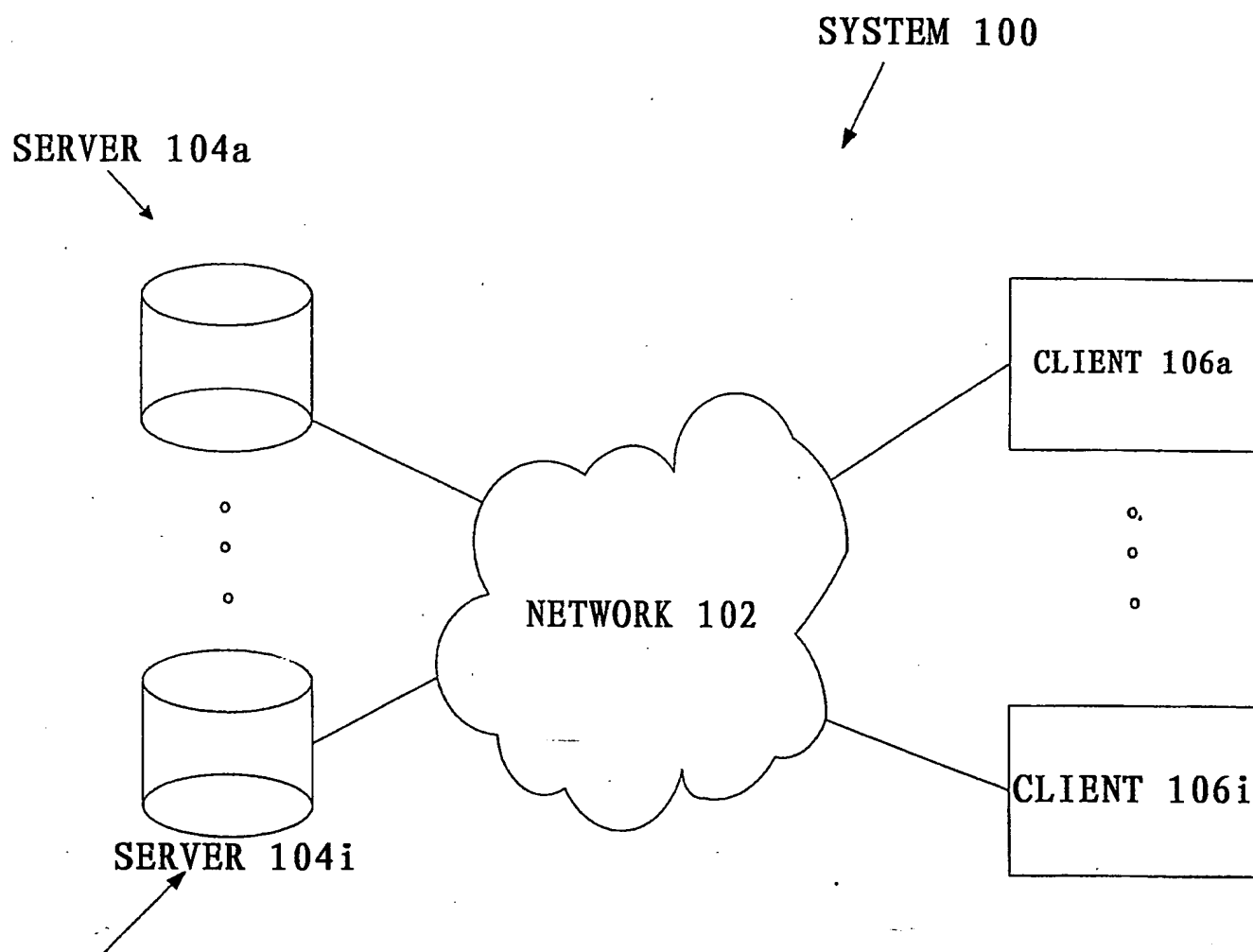


FIG. 1

2/7

SERVERS 104

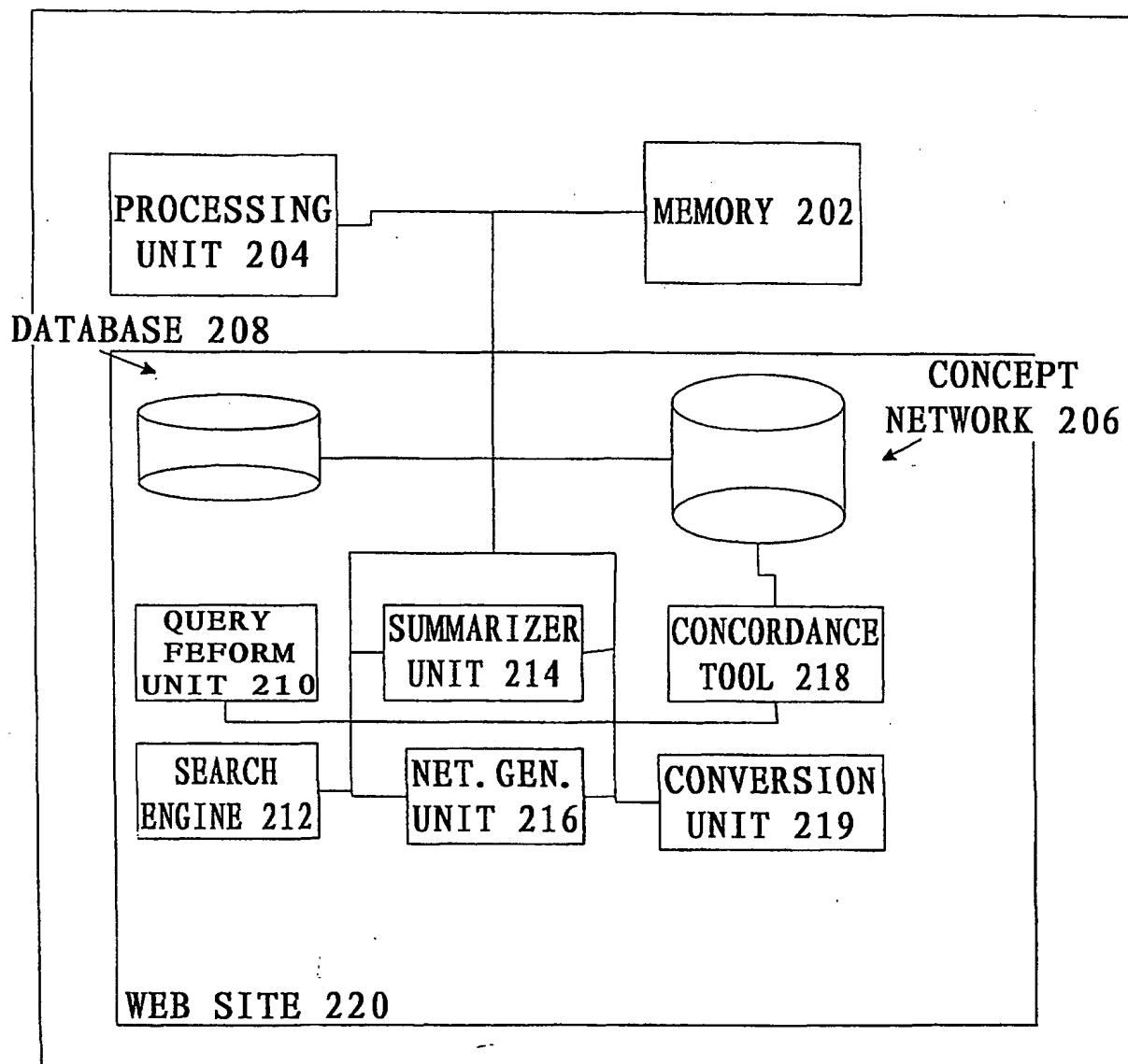
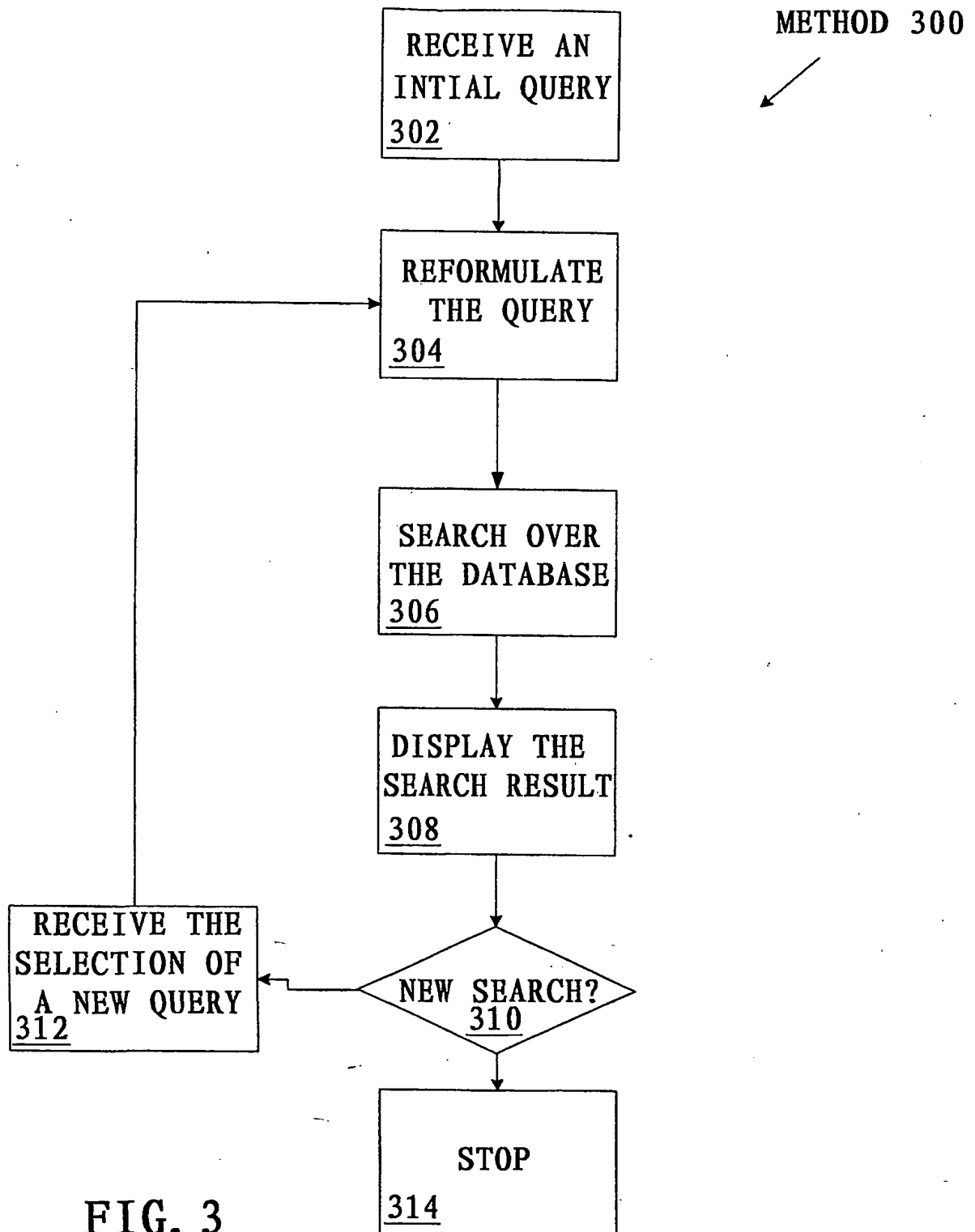


FIG. 2

3/7



4/7

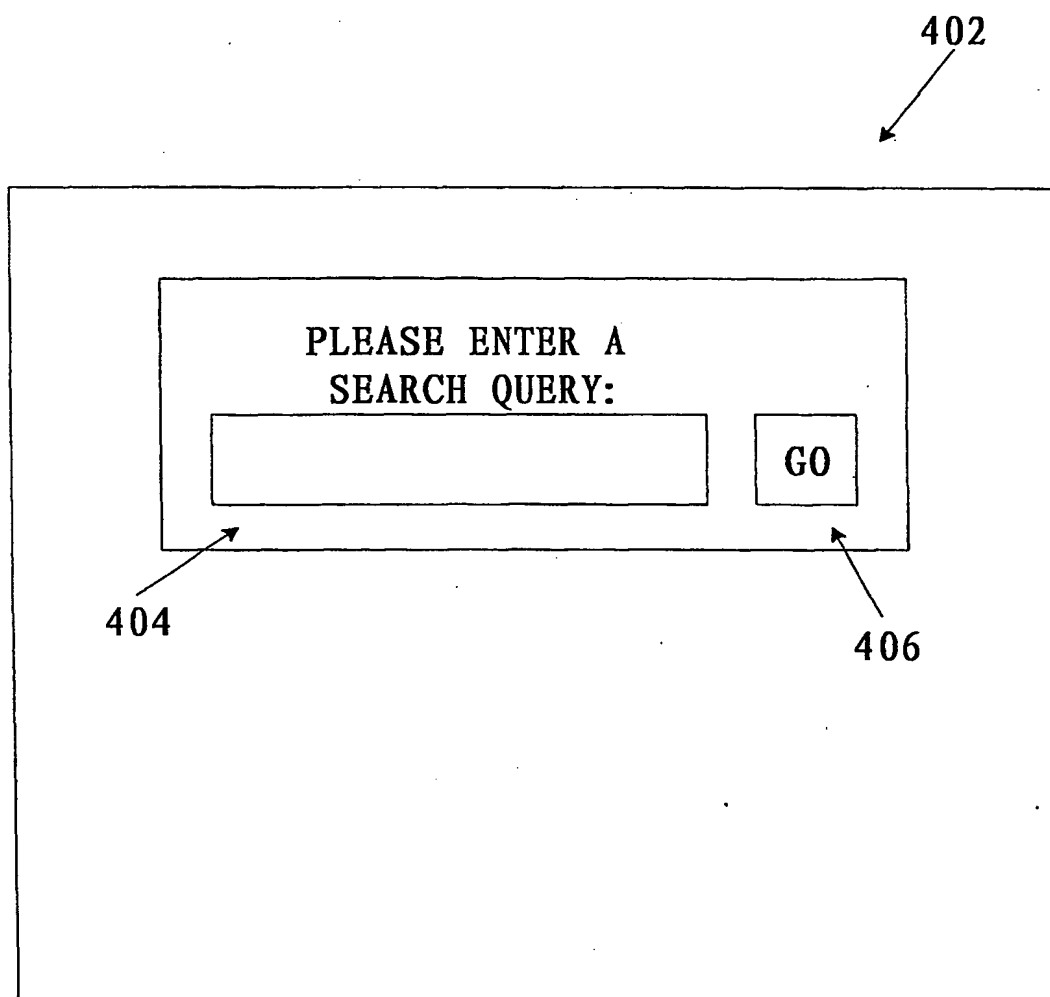


FIG. 4

5/7

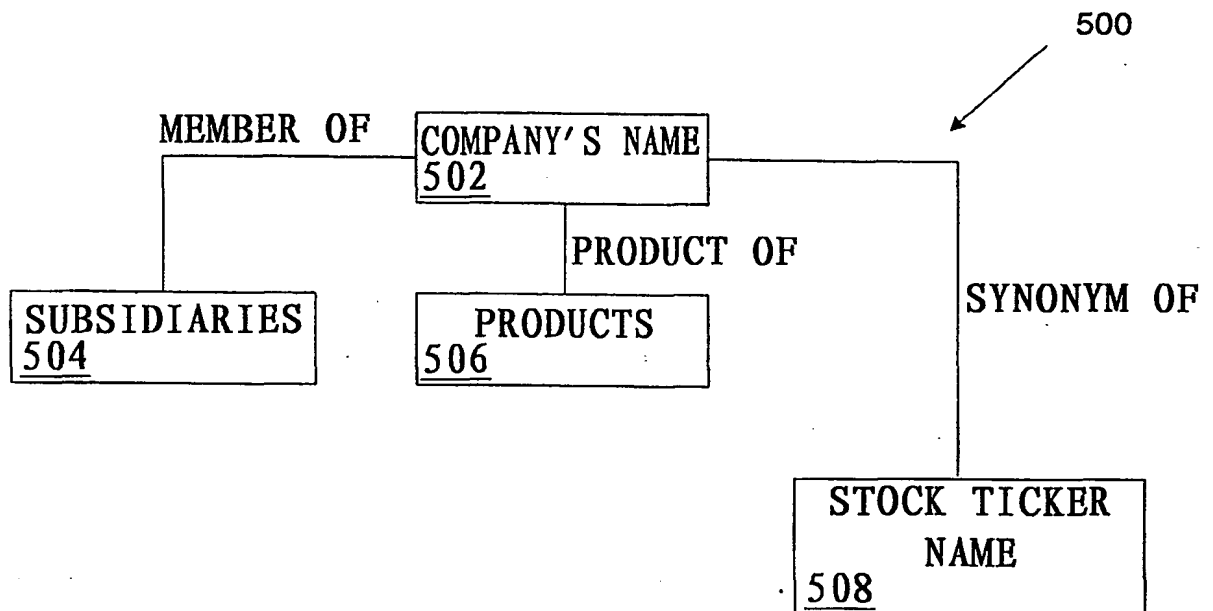


FIG. 5

6/7

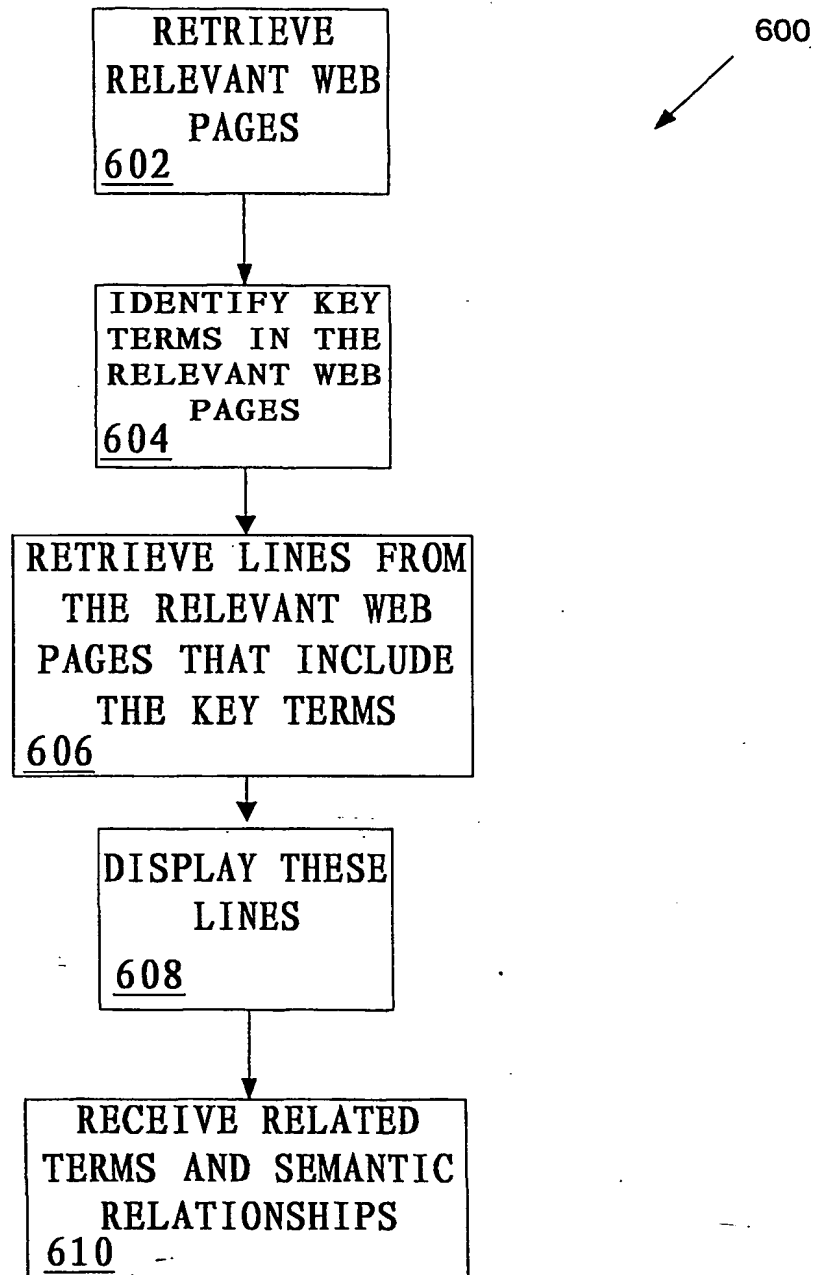


FIG. 6

7/7

DISPLAY 700

702b

702a

Baseball shoes include both rubber-soled and spikes.
The new commercial of the basketball shoes includes a superstar

o
o
o

702i

These dress shoes come in three different colors.

FIG. 7